BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Kongeriget Danmark

Patent application No.: 0744/97

Date of filing: 23 Jun 1997

Applicants: Sven Birkelund, Søtoften 26, 8250 Egå, DK;
Gunna Christiansen, Søtoften 26, 8250 Egå, DK

This is to certify the correctness of the following information:

The attached photocopy is a true copy of the following document:

- The specification, claims and drawings as filed with the application on the filing date indicated above.

Patentdirektoratet · Kongeriget Danmark ·

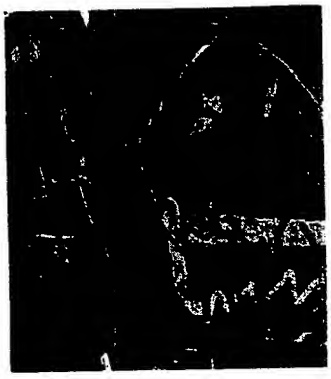
Erhvervsministeriet
Patentdirektoratet

TAASTRUP 17 Jul 1998

Lizzi Vester
Afdelingsleder

PLOUGMANN, VINGTOFT & PARTNERS

SANKT ANNÆ PLADS 11
POST OFFICE BOX 3007
DK - 1021 COPENHAGEN K
A/S REG. NO. 223795
TELEPHONE +45 33 63 93 00
TELEFAX +45 33 63 96 00
e-mail pv@pv.dk



Artist Bjørn Bjørnholt

23. juni 1997

Ny dansk patentansøgning
Sven Birkelund m.fl.
Novel surface exposed proteins from *Chlamydia pneumoniae*
Vor ref: 19922 DK 1

Plougmann, Vingtoft & Partners

FIELD OF THE INVENTION

The present invention relates to the identification of members of a gene family from the human respiratory pathogen *Chlamydia pneumoniae*, encoding surface exposed membrane proteins of a size of approximately 98/95 kDa. The invention relates to the novel DNA sequences, the deduced amino acid sequences of the corresponding proteins and the use of the DNA sequences and the proteins in diagnosis of infections caused by *C. pneumoniae*, in pathology, in epidemiology, and as vaccine components.

GENERAL BACKGROUND

C. pneumoniae is an obligate intracellular bacteria (Christiansen and Birkelund (1992); Grayston et al. (1986)). It has a cell wall structure as Gram negative bacteria with an outer membrane, a periplasmic space, and a cytoplasmic membrane. It is possible to purify the outer membrane from Gram negative bacteria with the detergent sarkosyl. This fraction is named the 'outer membrane complex (OMC)' (Caldwell et al. (1981)). The COMC (Chlamydia outer membrane complex) of *C. pneumoniae* contains four groups of proteins: A high molecular weight group of proteins 98/95 kDa, a double band of the cysteine rich outer membrane protein 2 (Omp2) protein of 62/60 kDa, the major outer membrane protein (MOMP) of 38 kDa, and the low-molecular weight lipo-protein Omp3 of 12 kDa. The Omp2/Omp3 and MOMP proteins are present in COMC from all Chlamydia species, and these genes have been cloned from both *C. trachomatis*, *C. psittaci* and *C. pneumoniae*. However, the genes encoding 98/95 kDa proteins from *C. pneumoniae* COMC have not been characterized or cloned.

The current state of *C. pneumoniae* serology and detection

C. pneumoniae is an obligate intra-cellular bacteria belonging to the genus Chlamydia which can be divided into four species: *C. trachomatis*, *C. pneumoniae*, *C. psittaci* and

C.pecorum. Common for the four species is their obligate intra cellular growth, and that they have a biphasic life cycle, with an extracellular infectious particle (the elementary body, EB), and an intercellular replicating form (the reticulate body, RB). In addition the Chlamydia species are characterized by a common lipopolysaccharide (LPS) epitope that is highly immunogenic in human infection. *C. trachomatis* is causing the human ocular infection (trachoma) and genital infections. *C. psittaci* is a variable group of animal pathogens where the avian strains can occasionally infect humans and give rise to a severe pneumonia (ornithosis). The first *C. pneumoniae* isolate was obtained from an eye infection, but it was classified as a non-typable Chlamydia. Under an epidemic outbreak of pneumonia in Finland it was realized that the patients had a positive reaction in the Chlamydia genus specific test, (the lygranum test), and the patients showed a titre increase to the untyped Chlamydia isolates. Similar isolates were obtained in an outbreak of upper respiratory tract infections in Seattle, and the Chlamydia isolates were classified as a new species, *Chlamydia pneumoniae* (Grayston et al. (1989)). In addition, *C. pneumoniae* is suggested to be involved in the development of atherosclerotic lesions and for initiating bronchial asthma (Kuo et al. (1995)). These two conditions are thought to be caused by either chronic infections, by a hypersensitivity reaction, or both.

Diagnosis of *Chlamydia pneumoniae* infections

Diagnosis of acute respiratory tract infection with *C. pneumoniae* is difficult. Cultivation of *C. pneumoniae* from patient samples is insensitive, even when proper tissue culture cells are selected for the isolation. A *C. pneumoniae* specific polymerase chain reaction (PCR) has been developed by Campbell et al.(1992) However, this PCR test can seldom detect *C. pneumoniae* DNA in patient samples because of the extremely low yield of these bacteria. Therefore, the only choice for detecting both acute and chronic infections is

sero-diagnostics. Sero-diagnosis of Chlamydia infections is currently based on either genus specific tests as the Lygranum test and ELISA, measuring the antibodies to LPS, or the more species specific tests where antibodies to purified EBs are measured by microimmuno fluorescence (Micro-IF) (Wang et al. (1970)). However, the micro-IF method is read by microscopy, and in order to ensure correct readings the result must be compared to the results with *C. trachomatis* used as antigen due to the cross-reacting antibodies to the common LPS epitope. Thus, there exists in the art an urgent need for development of reliable methods for species specific diagnosis of *Chlamydia pneumoniae*, as has been expressed in Kuo et al. (1995); "..a rapid reliable laboratory test of infection for the clinical laboratory is a major need in the field". Furthermore, the possible involvement of *C. pneumoniae* in atherosclerosis and bronchial asthma clearly warrants the development of an effective vaccine.

DETAILED DISCLOSURE OF THE INVENTION

The present invention aims at providing means for efficient diagnosis of infections with *Chlamydia pneumoniae* as well as the development of effective vaccines against infection with this microorganism. The invention thus relates to species specific diagnostic tests for infection in a mammal, such as a human, with *Chlamydia pneumoniae*, said tests being based on the detection of antibodies against surface exposed membrane proteins of a size of approximately 98/95 kDa, or the detection of nucleic acid fragments encoding such proteins or variants or subsequences thereof. The invention further relates to the amino acid sequences of proteins according to the invention, to variants and subsequences thereof, and to nucleic acid fragments encoding these proteins or variants or subsequences thereof. The present invention further relates to antibodies against proteins according to the invention. The invention also relates to the use of nucleic acid fragments and proteins according to the invention in

diagnosis of *Chlamydia pneumoniae* and vaccines against *Chlamydia pneumoniae*.

Prior to the disclosure of the present invention only a very limited number of genes from *C. pneumoniae* had been sequenced. These were primarily the genes encoding known *C. trachomatis* homologues: MOMP, Omp2, Omp3, Kdo-transferase, the heat shock protein genes GroEl/Es and DnaK, a ribonuclease P homologue and a gene encoding a 76 kDa protein of unknown function. The reason why so few genes have been cloned to date is the very low yield of *C. pneumoniae* which can be obtained after purification from the host cells. After such purification the DNA must be purified from the EBs, and at this step the *C. pneumoniae* DNA can easily be contaminated with host cell DNA. In addition to these inherent difficulties, it is exceedingly difficult to cultivate *C. pneumoniae* and use DNA technology to produce expression libraries with very low amounts (few μg) of DNA. It has been known since 1993 (Melgosa et al., 1993) that the 98/95 kDa protein is present in OMC from *C. pneumoniae*. However, only a very weak reaction with patient sera can be observed and prior to the work of the present inventors it has not been recognized that the 98/95 kDa proteins are surface exposed or that they are indeed immunogenic at all.

Use of antibodies to screen expression libraries is a well known method to clone fragments of genes encoding antigenic parts of proteins. However, since patient sera do not show a significant reaction with the 98/95 kDa proteins it has not been possible to use patient serum to clone the proteins.

It was known that monoclonal antibodies generated by the inventors reacted with conformational epitopes on the surface of *C. pneumoniae* and that they also reacted with *C. pneumoniae* OMC by immuno-electron microscopy (Christiansen et al. 1994). Furthermore, the 98/95 kDa proteins are the only unknown protein from the *C. pneumoniae* OMC (Christiansen et al. 1994). The inventors chose to take an unconventional step

in order to clone the hitherto 98/95 kDa proteins: *C. pneumoniae* OMC was purified and the highly immunogenic conformational epitopes were destroyed by SDS-treatment of the antigen before immunization. Thereby an antibody (PAG 150) to less immunogenic linear epitopes was obtained. This provided the possibility to obtain an antiserum which could detect the 98/95 kDa proteins in colony blotting of recombinant *E. coli*.

Mice infected with *C. pneumoniae* generate antibodies to the 98/95 kDa proteins identified by the inventors and named Omp4-7, but do not recognize the SDS treated heat denatured antigens normally used for SDS-PAGE and immunoblotting. However, a strong reaction was, seen when the antigen was not heat denatured. It is therefore highly likely that if a similar reaction is seen in connection with human infections the antigens of the present invention will be of invaluable use in sero-diagnostic tests and may very likely be used as a vaccine for the prevention of infections.

By generating antibodies against COMC from *C. pneumoniae* a polyclonal antibody (PAG 150) was obtained which reacted with all the proteins. This antibody was used to identify the genes encoding the 98/95 kDa proteins in an expression library of *C. pneumoniae* DNA. A problem in connection with the present invention was that a family comprising a number of similar genes were found in *C. pneumoniae*. Therefore, a large number of different clones were required to identify clusters of fragments. Only because the rabbit antibody generated by the use of SDS-denatured antigens contained antibodies to a high number of different epitopes positioned on different members of the protein family did the inventors succeed in cloning and sequencing four of the genes.The deduced amino acid sequences of two of the genes showed an amino acid homology of 61% and a size of 98 kDa, and these genes were fully sequenced except for approximately 45 base pairs in one of the genes. In addition, partial sequences from two other genes were obtained, from which the deduced

amino acid sequences showed an amino acid homology to the two fully sequenced genes in a similar range.

Part of the 98/95 kDa proteins were expressed as fusion proteins, and mice polyclonal monospecific antibodies against the proteins were produced. The antibodies reacted with the surface of *C. pneumoniae* in both immunofluorescence and immunoelectron microscopy. This shows for the first time that the 98/95 kDa protein family in *C. pneumoniae* comprises surface exposed outer membrane proteins. This important finding leads to the realization that members of the 98/95 kDa *C. pneumoniae* protein family are good candidates for the development of a sero diagnostic test for *C. pneumoniae*, as well as the development of a vaccine against infections with *C. pneumoniae* based on using these proteins. Furthermore, the proteins may be used as epidemiological markers, and polyclonal monospecific sera against the proteins can be used to detect *C. pneumoniae* in human tissue or detect *C. pneumoniae* isolates in tissue culture. Also, the genes encoding the 98/95 kDa protein family may be used for the development of a species specific diagnostic test based on nucleic acid detection/amplification.

A broad aspect of the present invention relates to a species specific diagnostic test for infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said test comprising detecting in a patient sample the presence of antibodies against proteins from the outer membrane of *Chlamydia pneumoniae*, said proteins being of a molecular weight of 98/95 kDa, or detecting the presence of nucleic acid fragments encoding said outer membrane proteins or fragments thereof.

In the context of the present application, the term "patient sample" should be taken to mean an amount of serum from a patient, such as a human patient, or an amount of plasma from said patient, or an amount of mucosa from said patient, or an amount of tissue from said patient, or an amount of urine

from said patient, or an amount of cerebrospinal fluid from said patient, or an amount of atherosclerotic lesion from said patient, or an amount of mucosal swaps from said patient, or an amount of cells from a tissue culture originating from said patient, or an amount of material which in any way originates from said patient.

In the present context, the term 98/95 kDa protein means proteins normally present in the outer membrane of *Chlamydia pneumoniae*, which in SDS-PAGE can be observed as one or more bands with an apparent molecular weight substantially in the range of 95-98 kDa.

Within the scope of the present invention are species specific sero-diagnostic tests based on the usage of the genes belonging to the gene family disclosed in the present application which might include the remaining sequences of the nucleic acid fragments disclosed herein which will be cloned using the approach described in detail in the examples below. Further using the sequences and antibodies already generated additional genes in the gene family of the invention will be sequenced by screening of existing genomic libraries generated by the inventors and subsequent sequencing of the inserts as described in the examples below.

Preferred embodiments of the present invention relate to species specific diagnostic tests according to the invention, wherein the outer membrane proteins have sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8, or variants or subsequences of these sequences.

When used in connection with proteins according to the present invention the term "variant" should be understood as a sequence of amino acids which shows a sequence homology of less than 100% to one of the proteins of the invention. A variant sequence can be of the same size or it can be of a different size as the sequence it is compared to. A variant

will typically show a sequence homology of at least 50%, preferably at least 60%, more preferably at least 70%, such as at least 80%, e.g. at least 90%, 95% or 98%.

The term "sequence homology" in connection with sequences of proteins of the invention means the percentage of matching amino acid residues (with respect to both position and type) in the proteins of the invention and an aligned protein of equal of different length.

Within the scope of the present invention are subsequences of one of the proteins of the invention, meaning a consecutive stretch of amino acid residues taken from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8. A subsequence will typically comprise at least 100 amino acids, preferably at least 80 amino acids, more preferably at least 70 amino acids, such as 50 amino acids. It might even be as small as 10-50 amino acids, such as 20-40 amino acids, e.g. about 30 amino acids.A subsequence will typically show a sequence homology of at least 50%, preferably at least 60%, more preferably at least 70%, such as at least 80%, e.g. at least 90%, 95% or 98%.

Diagnostic tests according to the invention include immunoassays selected from the group consisting of a direct or indirect EIA such as an ELISA, an immunoblot technique such as a Western blot, a radio immuno assay, and any other non-enzyme linked antibody binding assay or procedure such as a fluorescence, agglutination or precipitation reaction, and nephelometry.

A preferred embodiment of the present invention relates to species specific diagnostic tests according to the invention, said test comprising an ELISA, wherein antibodies against the proteins of the invention or fragments thereof are detected in samples.

A preferred embodiment of the invention, is an ELISA based on detection in samples of antibodies against proteins of the invention. The ELISA may use proteins of the invention, or variants thereof, i.e. the antigen, as coating agent. An ELISA will typically be developed according to standard methods well known in the art, such as methods described in "Antibodies; a laboratory manual", Ed. David Lane Harlow, Cold Spring Habor laboratories (1988), which is hereby incorporated by reference.

Recombinant proteins will be produced using DNA sequences obtained essentially using methods described in the examples below. Such DNA sequences, comprising the entire coding region of each gene in the gene family of the invention, will be cloned into an expression vector from which the deduced protein sequence can be purified. The purified proteins will be analyzed for reactivity in ELISA using both monoclonal and polyclonal antibodies as well as sera from experimentally infected mice and human patient sera.

From the experimentally infected mice sera it is known that non-linear epitopes are recognized predominantly. Thus, it is contemplated that different forms of purification schemes known in the art will be used to analyze for the presence of discontinuous epitopes, and to analyze whether the human immune response is also directed against such epitopes.

Preferred embodiments of the present invention relate to species specific diagnostic tests according to the invention, wherein the nucleic acid fragments have sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 and SEQ ID NO: 7, or variants of these sequences.

In connection with nucleic acid fragments according to the present invention the term "variant" should be understood as a sequence of nucleic acids which shows a sequence homology of less than 100%. A variant sequence can be of the same size or it can be of a different size as the sequence it is

compared to. A variant will typically show a sequence homology of at least 50%, preferably at least 60%, more preferably at least 70%, such as at least 80%, e.g. at least 90%, 95% or 98%.

The term "sequence homology" in connection with nucleic acid fragments of the invention means the percentage of matching nucleic acids (with respect to both position and type) in the nucleic acid fragments of the invention and an aligned nucleic acid fragment of equal or different length.

In order to obtain information concerning the general distribution of each of the genes according to the present invention, PCR will be performed for each gene on all available *C. pneumoniae* isolates. This will provide information on the general variability of the genes or nucleic acid fragments of the invention. Variable regions will be sequenced. From patient samples PCR will be used to amplify variable parts of the genes for epidemiology. Non-variable parts will be used for amplification by PCR and analyzed for possible use as a diagnostic test. It is contemplated that if variability is discovered, PCR of variable regions can be used for epidemiology. PCR of non-variable regions can be used as a species specific diagnostic test. Using genes encoding proteins known to be invariable in all known isolates prepared as targets for PCR to genes encoding proteins with unknown function.

Particularly preferred embodiments of the present invention, relate to diagnostic tests according to the invention, wherein detection of nucleic acid fragments is obtained by using nucleic acid amplification, preferably polymerase chain reaction (PCR).

Within the scope of the present invention is a PCR based test directed at detecting nucleic acid fragments of the invention or variants thereof. A PCR test will typically be developed according to methods well known in the art and will typically

comprise a PCR test capable of detecting and differentiating between nucleic acid fragments of the invention. Preferred are quantitative competitive PCR tests or nested PCR tests. The PCR test according to the invention will typically be developed according to methods described in detail in EP B 540 588, EP A 586 112, EP A 643 140 OR EP A 669 401, which are hereby incorporated by reference.

Within the scope of the present invention are variants and subsequences of one of the nucleic acid fragments of the invention, meaning a consecutive stretch of nucleic acids taken from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7. A variant or subsequence will preferably comprise at least 100 nucleic acids, preferably at least 80 nucleic acids, more preferably at least 70 nucleic acids, such as at least 50 nucleic acids. It might even be as small as 10-50 nucleic acids, such as 20-40 nucleic acids, e.g. about 30 nucleic acids. A subsequence will typically show a sequence homology of at least 50%, preferably at least 60%, more preferably at least 70%, such as at least 80%, e.g. at least 90%, 95% or 98%.

A very important aspect of the present invention relates to proteins of the invention derived from *Chlamydia pneumoniae* having amino acid sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences thereof having a sequence homology of at least 50%, preferably at least 60%, more preferably at least 70%, such as at least 80%, e.g. at least 90%, 95% or 98% and a substantially identical biological activity.

By the term "substantially identical biological activity" is meant that a protein when used in immunization and immunoblotting, as described in detail in Example 2 herein, will yield a reaction in serum from a patient, such as a human patient, infected with *Chlamydia pneumoniae* which is at least 50% of the reaction obtained with the purified fusion

protein transcribed from the pEX1-1 clone, as described herein, preferably at least 60% of said reaction, more preferably at least 70 % of said reaction, still more preferably 80% of said reaction, even more preferably at least 90% of said reaction.

Comparison of the DNA sequences from genes encoding Omp4-7 shows that the overall homology between the individual genes ranges between approximately 46-53%. Comparison of the amino acid sequences of Omp4-7 shows that the overall homology of the amino acids ranges between 48-63%. The homology is generally maintained along the entire length of the deduced amino acids. However, as seen from figure 12 there are some regions in which the homology is more pronounced. This is seen in the repeated sequence where the sequence GGAI is repeated six times in two of the genes. In the third gene this part of the sequence is not present as an indication of a deletion of this part of the gene. For the fourth gene, the sequence encoding the repeated part has not yet been cloned and sequenced. It is interesting that the DNA homology is not conserved for the sequences encoding the four amino acids GGAI. This may indicate a functional role of this part of the protein and indicates that the repeated structure did not occur by a duplication of the gene. In addition to the four amino acid repeats GGAI a region from amino acid 400 to 490 has a higher degree of homology than the rest of the protein, with the conserved sequence FSGE occurring in all sequences.

Since none of the genes and deduced amino acid sequences of the invention are identical the following is within the scope of the present invention; production of monospecific antibodies, the use of said antibodies for characterizing which *C. pneumoniae* proteins are expressed, the use of said antibodies for characterizing at which time during developmental life cycle said *C. pneumoniae* proteins are expressed, and the use of said antibodies for characterizing the precise cellular localization of said *C. pneumoniae* proteins. Also within the scope of the present invention is

the use of monospecific antibodies against proteins of the invention for determining which part of said proteins is surface exposed and how proteins in the *C. pneumoniae* COMC interact with each other.

Preferred embodiments of the present invention relate to polypeptides which comprise subsequences of the proteins of the invention, said subsequences comprising the sequence GGAI. Further preferred embodiments of the present invention relate to polypeptides which comprise subsequences of the proteins of the invention, said subsequences comprising the sequence FSGE

Polypeptides according to the invention will typically be of a length of at least 10 amino acids, preferably at least 15 amino acids, preferably at least 20 amino acids, preferably at least 25 amino acids, preferably at least 30 amino acids, preferably at least 35 amino acids, preferably at least 40 amino acids, preferably at least 45 amino acids, preferably at least 50 amino acids, preferably at least 55 amino acids, preferably at least 100 amino acids.

A very important aspect of the present invention relates to nucleic acid fragments of the invention derived from *Chlamydia pneumoniae*, variants and subsequences thereof

A further important aspect of the present invention relates to at least partially purified nucleic acid fragments according to the invention.

Another important aspect of the present invention relates to antibodies against the proteins according to the invention, such antibodies including polyclonal monospecific antibodies and monoclonal antibodies against proteins with sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences of these sequences.

A very important aspect of the present invention relates to diagnostic kits for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kits comprising one or more proteins with amino acid sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences of these sequences.

Another very important aspect of the present invention relates to diagnostic kits for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kits comprising antibodies against a protein with an amino acid sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or subsequences or variants of these sequences. Antibodies included in a diagnostic kit according to the invention can be polyclonal or monoclonal or a mixture hereof.

Still another very important aspect of the present invention relates to diagnostic kits for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kits comprising one or more nucleic acid fragments with sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 and SEQ ID NO: 7 or variants or subsequences of these sequences.

An aspect of the present invention relates to a composition for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*, said composition comprising one or more proteins with amino acid sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences of these sequences.

An important role for the proteins of the invention in prevention of infection of a mammal, such as a human, with *C. pneumoniae* is expected. Thus proteins of the invention, including variants and subsequences will be produced, typically by using recombinant techniques, and will then be

used as an antigen in immunization of mammals, such as rabbits. Subsequently, the hyper immune sera obtained by the immunization will be analyzed for protection against *C. pneumoniae* infection using a tissue culture assay. In addition it is contemplated that monoclonal antibodies will be produced, typically using standard hybridoma techniques, and analyzed for protection against infection with *C. pneumoniae*.

It is envisioned that particularly interesting and immunogenic epitopes will be found in connection with the proteins of the invention, which will comprise subsequences of said proteins. It is preferred to use polypeptides comprising such subsequences of the proteins of the invention in immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

An important aspect of the present invention relates to the use of proteins with sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences thereof in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

A preferred embodiment of the present invention relates to the use of proteins according to the invention in an undenatured form, in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

A very important aspect of the present invention relates to the use of proteins with sequences selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 and SEQ ID NO: 8 or variants or subsequences of these sequences, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

A preferred embodiment of the present invention relates to the use of proteins according to the invention in an

undenatured form, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

A very important aspect of the present invention relates to the use of nucleic acid fragments with nucleotide sequences selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 and SEQ ID NO: 7 or variants or subsequences of said nucleotide sequences for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

It is envisioned that one type of vaccine against *C. pneumoniae* will be developed by using gene-gun vaccination of mice. Typically, different genetic constructs containing nucleic acid fragments, combinations of nucleic acid fragments according to the invention will be used in the gene-gun approach. The mice will then subsequently be analyzed for production of both humoral and cellular immune response and for protection against infection with *C. pneumoniae* after challenge herewith.

The invention therefore also relates to a method for immunizing an mammal, such as a human, with outer membrane protein, according to the invention, derived from *Chlamydia pneumoniae*, the method comprising administering an immunogenically effective amount of said outer membrane protein.

In line with this, the invention also relates to the uses of the proteins of the invention as a pharmaceutical (a vaccine) as well as to the uses thereof for the preparation of a vaccine against infections with *Chlamydia pneumoniae*.

Preparation of vaccines which contain protein sequences as active ingredients is generally well understood in the art, as exemplified by U.S. Patents 4,608,251; 4,601,903; 4,599,231; 4,599,230; 4,596,792; and 4,578,770, all incorporated herein by reference. Typically, such vaccines are prepared as injectables either as liquid solutions or suspen-

sions; solid forms suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified. The active immunogenic ingredient is often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like, and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, or adjuvants which enhance the effectiveness of the vaccines.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkalene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1-2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10-95% of active ingredient, preferably 25-70%.

The protein sequences may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides,

and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective and immunogenic. The quantity to be administered depends on the subject to be treated, including, e.g., the capacity of the individual's immune system to mount an immune response, and the degree of protection desired. Suitable dosage ranges are of the order of several hundred micrograms active ingredient per vaccination with a preferred range from about 0.1 μg to 1000 μg, such as in the range from about 1 μg to 300 μg, and especially in the range from about 10 μg to 50 μg. Suitable regimes for initial administration and booster shots are also variable but are typified by an initial administration followed by subsequent inoculations or other administrations.

The manner of application may be varied widely. Any of the conventional methods for administration of a vaccine are applicable. These are believed to include oral application on a solid physiologically acceptable base or in a physiologically acceptable dispersion, parenterally, by injection or the like. The dosage of the vaccine will depend on the route of administration and will vary according to the age of the person to be vaccinated and, to a lesser degree, the size of the person to be vaccinated.

Some of the proteins of the vaccine are expected to be sufficiently immunogenic in a vaccine, but for some of the others the immune response may be enhanced if the vaccine further comprises an adjuvant substance.

Various methods of achieving adjuvant effect for the vaccine include use of agents such as aluminum hydroxide or phosphate (alum), commonly used as 0.05 to 0.1 percent solution in phosphate buffered saline, admixture with synthetic polymers of sugars (Carbopol) used as 0.25 percent solution, aggrega-

tion of the protein in the vaccine by heat treatment with temperatures ranging between 70° to 101°C for 30 second to 2 minute periods respectively. Aggregation by reactivating with pepsin treated (Fab) antibodies to albumin, mixture with bacterial cells such as *C. parvum* or endotoxins or lipopolysaccharide components of gram-negative bacteria, emulsion in physiologically acceptable oil vehicles such as mannide monooleate (Aracel A) or emulsion with 20 percent solution of a perfluorocarbon (Fluosol-DA) used as a block substitute may also be employed. Other interesting candidates for adjuvants are DDA (dimethyldioctadecylammonium bromide), but also Freund's complete and incomplete adjuvants as well as QuilA and RIBI are interesting possibilities.

Other possibilities involve the use of immunomodulating substances such as lymphokines (e.g. IFN-γ, IL-2 and IL-12) or synthetic IFN-γ inducers such as poly I:C in combination with the above-mentioned adjuvants.

In many instances, it will be necessary to have multiple administrations of the vaccine, usually not exceeding six vaccinations, more usually not exceeding four vaccinations and preferably one or more, usually at least about three vaccinations. The vaccinations will normally be at from two to twelve week intervals, more usually from three to five week intervals. Periodic boosters at intervals of 1-5 years, usually three years, will be desirable to maintain the desired levels of protective immunity. The course of the immunization may be followed by *in vitro* assays. The assays may be performed using conventional labels, such as radionuclides, enzymes, fluorescers, and the like. These techniques are well known and may be found in a wide variety of patents, such as U.S. Patent Nos. 3,791,932; 4,174,384 and 3,949,064, as illustrative of these types of assays.

It is also possible to produce a living vaccine by introducing, into a non-pathogenic microorganism, at least one nucleic acid fragment encoding a protein fragment or protein

of the invention, and effecting expression of the protein fragment or the protein on the surface of the microorganism (e.g. in the form of a fusion protein including a membrane anchoring part or in the form of a slightly modified protein or protein fragment carrying a lipidation signal which allows anchoring in the membrane). The skilled person will know how to adapt relevant expression systems for this purpose

Another part of the invention is based on the fact that recent research have revealed that a DNA fragment cloned in a vector which is non-replicative in eukaryotic cells may be introduced into an animal (including a human being) by e.g. intramuscular injection or percutaneous administration (the so-called "gene gun" approach). The DNA is taken up by e.g. muscle cells and the gene of interest is expressed by a promoter which is functioning in eukaryotes, e.g. a viral promoter, and the gene product thereafter stimulates the immune system. These newly discovered methods are reviewed in Ulmer et al., 1993, which hereby is included by reference.

Thus, a nucleic acid fragment encoding a protein or protein of the invention may be used for effecting *in vivo* expression of antigens, *i.e.* the nucleic acid fragments may be used in so-called DNA vaccines. Hence, the invention also relates to a vaccine comprising a nucleic acid fragment encoding a protein fragment or a protein of the invention, the vaccine effecting *in vivo* expression of antigen by an mammal, such as a human, to whom the vaccine has been administered, the amount of expressed antigen being effective to confer substantially increased resistance to infections with *Chlamydia pneumoniae* in an mammal, such as a human.

The efficacy of such a "DNA vaccine" can possibly be enhanced by administering the gene encoding the expression product together with a DNA fragment encoding a protein which has the capability of modulating an immune response. For instance, a gene encoding lymphokine precursors or lymphokines (e.g. IFN-γ, IL-2, or IL-12) could be administered together with the

gene encoding the immunogenic protein fragment or protein, either by administering two separate DNA fragments or by administering both DNA fragments included in the same vector. It is also a possibility to administer DNA fragments comprising a multitude of nucleotide sequences which each encode relevant epitopes of the protein fragments and proteins disclosed herein so as to effect a continuous sensitization of the immune system with a broad spectrum of these epitopes.

The following experimental non-limiting examples are intended to illustrate certain features and embodiments of the invention.

LEGENDS TO FIGURES

Figure 1. The figure shows electron microscopy of negative stained purified *C. pneumoniae* EB (A) and purified OMC (B).

Figure 2. The figure shows silver stained 15% SDS-PAGE of purified EB and OMC. Lane 1, purified *C. pneumoniae* EB; lane 2, *C. pneumoniae* OMC; lane 3, purified *C. trachomatis* EB; and lane 4 *C. trachomatis* OMC.

Figure 3. The figure shows immunoblotting of *C. pneumoniae* EB separated by 10% SDS-PAGE, transferred to nitrocellulose and reacted with rabbit anti *C. pneumoniae* OMC.

Figure 4. The figure shows coomassie blue stained 7.5% SDS-PAGE of recombinant pEX that were detected by the rabbit anti *C. pneumoniae* serum. Arrow indicated the localization of the 117 kDa b-galactosidase protein.

Figure 5. The figure shows immunoblotting of recombinant pEX colones detected by colony blotting separated by 7.5% SDS-PAGE and transferred to nitrocellulose and reacted with rabbit anti *C. pneumoniae* OMC. Lane 1, seablue molecular weight standard. Lane 2-6 pEX clones cultivated at 42°C to

induce the production of the b-galactosidase fusion proteins.

Figure 6. The figure shows assembled pEX contics.

Figure 7. The figure shows sequence strategy for omp4 and omp5. Arrows indicates primers used for sequencing.

Figure 8. The figure shows alignment of *C. pneumoniae* Omp4 and Omp5, using the program pileup in the GCG package.

Figure 9. The figure shows immunofluorescence of *C. pneumoniae* infected HeLa, 72 hrs. after infection, reacted with mouse monospecific anti-serum against pEX3-36 fusion protein. pEX3-36 is a part of the omp5 gene.

Figure 10. The figure shows immunoblotting of *C. pneumoniae* EB, lane 1-3 heated to 100°C in SDS-sample buffer, lane 4-6 unheated. Lane 1 reacted with rabbit anti *C. pneumoniae* OMC; lane 2 and 4 pre-serum; lane 3 and 5 polyclonal rabbit anti pEX1-1 fusion protein; lane 6 MAb 26.1.

Figure 11. The figure shows immunoblotting of *C. pneumoniae* EB, lane 1-4 heated to 100oC in SDS-sample buffer, lane 5-6 unheated. Reacted with serum from C57-black mice 14 days after infection with 10^7 CFU of *C. pneumoniae*. Lane 1 and 5 mouse 1; lane 2 and 6 mouse 2; lane 3 and 5 mouse 3; and lane 4 and 8 mouse 4.

Figure 12. The figure shows alignment of *C. pneumoniae* Omp4, Omp5, Omp6 and Omp7 using the program pileup in the GCG package.

EXAMPLE 1

Cloning of the genes encoding the 98/95 kDa *C. pneumoniae* COMC proteins

Purification of *C. pneumonia* EBs and COMC

C. pneumoniae was cultivated in HeLa cells. Cultivation was done according to the specifications of Miyashita and Matsumoto (1992), with the modification that centrifugation of supernatant and of the later precipitate and turbid bottom layer was carried out at 100,000 X g. The microorganism attached to the HeLa cells by 30 minutes of centrifugation at 1000 x g, after which the cells were incubated in RPMI 1640 medium (Gibco BRL, Germany cat No. 51800-27), containing 5% foetal calf serum (FCS, Gibco BRL, Germany Cat No. 10106.169) gentamicin for two hours at 37°C in 5% CO2 atmosphere. The medium was changed to medium that in addition contained 1 mg per ml of cycloheximide. After 48 hours of incubation a coverslip was removed from the cultures and the inclusion was tested with an antibody specific for *C. pneumoniae* (MAb 26.1) (Christiansen et al. 1994) and a monoclonal antibody specific for the species *C. trachomatis* (MAb 32.3, Loke diagnostics, Århus Denmark) to ensure that no contamination with *C. trachomatis* had occurred. The HeLa cells were tested by Hoechst stain for Mycoplasma contamination as well as by culture in BEa and BEg medium (Freund et al., 1979). Also the *C. pneumoniae* stocks were also tested for Mycoplasma contamination by cultivation in BEa and BEg medium. No contamination with *C. trachomatis*, Mycoplasmas or bacteria were detected in cultures or cells. 72 hours post-infection the monolayer was washed in PBS, the cells were loosened in PBS with a rubber policeman, and the Chlamydia were liberated from the host cell by sonication. The *C. pneumoniae* EBs and RBs were purified on discontinuous density gradients (Miyashita et al. (1992)). The purity of the Chlamydia EBs were verified by negative staining and electronmicroscopy (Figure 1), only particles of a size of 0.3 to 0.5 mm were

detected in agreement with the structure of *C. pneumonia* EBs. The purified Chlamydia EBs were subjected to sarkosyl extraction as described by Caldwell et al (1981) with the modification that a brief sonication was used to suspend the COMC. The purified COMC was tested by electronmicroscopy and negative staining (Figure 1), where a folded outer membrane complex was seen.

SDS-PAGE analysis of purified EBs and COMC

The proteins from purified EBs and *C. pneumoniae* OMC were separated on 15% SDS-polyacrylamide gel, and the gel was silver stained (Figure 2), in lane 1 it is seen that the purified EBs contain major proteins of 98/95 kDa and a protein of 38 kDa, in the purified COMC (lane 2) these two protein groups are also dominant. In addition, proteins with a molecular weight of 62/60 kDa, 55 kDa, and 12 kDa have been enriched in the COMC preparation. When the purified *C. pneumoniae* EBs are compared to purified *C. trachomatis* EB (lane 3) it is seen that predominant protein in the *C. trachomatis* EB is the major outer membrane protein (MOMP), and it is also the dominant band in the COMC preparation of *C. trachomatis* (lane 4), and Omp2 of 60/62 kDa as well as Omp3 at 12 kDa are seen in the preparation. However, no major bands with a size of 98/95 kDa are detected as in the *C. pneumoniae* COMC preparation.

Production of rabbit polyclonal antibodies against *C. pneumoniae* COMC

To ensure production of rabbit antibodies that would recognize all the *C. pneumoniae* proteins in immuno-blotting and colony-blotting 10 µg of COMC antigen was dissolved in 20 µl of SDS sample buffer and thereafter divided into 5 vials. The dissolved antigen was further diluted in one ml of PBS and one ml of Freund incomplete adjuvant (Difco laboratories, USA cat. No. 0639-60-6) and injected into the quadrisceps muscle of a New Zealand white rabbit. The rabbit was given

three times intramuscular injections at an interval of one week, and after further three weeks the dissolved COMC protein, diluted in one ml PBS was injected intravenously, and the procedure was repeated two weeks later. Eleven weeks after the beginning of the immunization, the serum was obtained from the rabbit. Purified *C. pneumoniae* EBs were separated by SDS-PAGE, and the proteins were electrotransferred to nitrocellulose membrane. The membrane was blocked and immunostained with the polyclonal COMC antibody (Figure 3). The serum recognized proteins with a size of 98/95, 60 and 38 kDa in the EB preparation. This is in agreement with the sizes of the outer membrane proteins.

Cloning of the COMC proteins

Due to the cultivation of *C. pneumoniae* in HeLa cells, contaminating host cell DNA could be present in the EB preparations. Therefore, the purified EB preparations were treated with DNAse to remove contaminating DNA. The *C. pneumoniae* DNA was then purified by CsCl gradient centrifugation. The *C. pneumoniae* DNA was partially digested with Sau3A and the fractions containing DNA fragments with a size of approx. 0.5 to 4.0 kb were cloned into the expression vector system pEX (Boehringer, Germany cat. No. 1034 766, 1034 774, 1034 782). The pEX vector system has a β-galactosidase gene with multiple cloning sites in the 3'end of the β-galactosidase gene. Expression of the gene is regulated by the PR promoter, so the protein expression can be induced by elevating the temperature from 32 to 42°C. The colonies of recombinant bacteria were transferred to nitrocellulose membranes, and the temperature was increased to 42°C for two hours. The bacteria were lysed by placing the nitrocellulose membranes on filters soaked in 5% SDS. The colonies expressing outer membrane proteins were detected with the polyclonal antibody raised against *C. pneumoniae* COMC. The positive clones were cultivated in suspension and induced at 42°C for two hours. The protein profile of the clones were analysed by SDS-PAGE, and increases in the size

of the induced b-galactosidase were observed (Figure 4). In addition, the proteins were electrotransferred to nitrocellulose membranes, and the reaction with the polyclonal serum against COMC was confirmed (Figure 5).

Sequencing of positive COMC clones

To characterize the pEX clones, the inserted *C. pneumoniae* DNA was sequenced. The resulting DNA sequences were searched against the prokaryotic sequences in the GenEmbl database. The search identified 6 clones as part of the omp2 gene, and 2 clones as part of the omp3 gene, and 2 clones as part of the MOMP gene, indicating that COMC proteins had been successfully cloned. Furthermore, 32 clones were obtained, containing DNA sequences not found in the GenEmbl database. These sequences could, however, be clustered in two contics of 6 and 4 clones, and three clones were identical. In addition 19 clones were found with no overlap to the contics (Figure 6). To obtain more sequence data for the genes, *C. pneumoniae* DNA was totally digested with BamHI restriction enzyme, and the fragments were cloned into the vector pBluescript. The ligated DNA was electrotransformed into *E. coli* XL1-Blue and selected on plates containing Ampicillin. The recombinant bacterial colonies were transferred to a nitrocellulose membrane, and colony hybridisation was performed using the inserts of pEX 1-1 clone as a probe. A clone containing a single BamHI fragment of 4.5 kb was found, and the hybridisation to the probe was confirmed by Southern blotting. The insert of the clone was sequenced bi-directionally using synthetic primers for approx. each 300 bp. The sequence of the *Bam*HI fragment made it possible to join the two contics of pEX clones. Totally, together with the pEX clones it was possible to assemble 6.5 kb DNA sequence, encoding two new COMC proteins (Figure 7).

Analysis of DNA sequence

The 6.5 kb DNA sequence encoded 2 proteins with a size of 98 kDa. The genes were transcribed in opposite directions. The genes were named omp4 and omp5 (Figure 7). Downstream omp4 a possible termination structure was located. The 3'end of the omp5 gene was not cloned due to the presence of the BamHI restriction enzyme site positioned within the gene. The translated DNA sequence of omp4 and omp5 was compared by use of the gap programme in the GCG package (Wisconsin package, version 8.1-UNIX, August 1995, sequence analysis software package) (Figure 8). The two genes had an amino acid identity of 41% (homology 61%), and a possible cleavage site for signal peptidase 1 was present at amino acid 17 in Omp4 and amino acid 25 in Omp5. When the amino acid sequence encoded by two other pEX clones were compared to the sequence of omp4 and omp5 they also had amino acid homology to the genes. It is seen that the two clones have homology to the same area in the omp4 and omp5 proteins. Consequently, the pEX clones must have originated from two additional genes. Therefore these genes were named omp6 and omp7. Multiple alignment of the four genes showed in addition that omp6 had a deletion in the sequence. The area deleted in omp6 is a sequence coding for 225 amino acids. A sequence of the amino acids GGAI (Gly, Gly, Ala, Ile) was present 6 times in the sequence. These repeats were conserved in both omp4 and omp5.

EXAMPLE 2

Polyclonal monospecific antibodies against pEX fusion proteins

To investigate the topology of the Omp4-7 proteins, representative pEX clones, were selected from each gene. The fusion proteins of β-galactosidase/Omp were induced, and the proteins were partially purified as inclusion bodies. Balb/c mice were immunized three times intramuscular with the antigens at an interval of one week, and after six weeks the

serum was obtained from the mice. HeLa cells were infected with the *C. pneumoniae*. 72 hours after the infection the mono-layers were fixed with 3.7% formaldehyde. This treatment makes the outer membrane of the Chlamydia impermeable for antibodies due to the extensive cross-linking of the outer membrane proteins by the formaldehyde. The HeLa cells were permeabilized with 0.2% Triton X100, the monolayers were washed in PBS, then incubated with 20% (v/v) FCS to inactivate free radicals of the formaldehyde. The mice sera were diluted 1:100 PBS with 20% (v/v) FCS and incubated with the monolayers for half an hour. The monolayers were washed in PBS and secondary FITCH conjugated rabbit anti mouse serum was added for half an hour, and the monolayers were washed and mounted. Several of the antibodies reacted strongly with the EBs in the inclusions (Figure 9). In spite of the formaldehyde fixation it could not be excluded that the surface of the EB was changed by the treatments, so that the antibodies could get access to the Omp4-7. Therefore, the reaction was confirmed by immuno-electron microscopy with the antibody raised against clone pEX3-36. Purified EB of *C. pneumoniae* were absorbed to carbon coated nickel grids. After the absorption the grids were washed with PBS and blocked in 0.5% Ovalbumin dissolved in PBS. The antibodies were diluted 1:100 in the same buffer and incubated for 30 minutes. The grids were washed in PBS. Rabbit anti mouse Ig conjugated with 10nm colloidal gold diluted in PBS containing 1% gelatin was added to the grids for half an hour. The grids were washed in 3 x PBS with 1% gelatin and 3 times in PBS, the grids were contrastained with 0.7% phospho tungstic acid. The grids were analysed in a Jeol 1010 electron microscope at 40 kV. It was seen that the gold particles were covering the surface of the purified EB. Because the *C. pneumoniae* EBs were not exposed to any detergent or fixation under either the purification or the reaction with antibodies, these results show that the cloned proteins have surface exposed epitopes.

Immuno blotting analysis with hyperimmune monospecific rabbit anti-serum

The insert of pEX1-1 clone was amplified by PCR using primers containing LIC sites. The PCR product could therefore be inserted in the pET-32 LIC vector (Novagen, UK cat No. 69076-1). Thereby the insert sequence of the pEX1-1 clone was expressed in the new vector as a fusion protein, the part of the fusion protein encoded by the pET-32 LIC vector had 6 histidin residues in a row. The expression of the fusion protein was induced in this vector, and the fusion protein could be purified under denaturing condition on a Ni2+ column due to the high affinity of the histidine residues to divalent cations. The purified protein was used for immunization of a New Zealand white rabbit. After 6 times intramuscular and 2 times intravenous immunization the serum was obtained from the rabbit. Purified *C. pneumoniae* EB was dissolved in SDS-sample buffer. Half of the sample was heated to 100°C in the sample buffer, whereas the other half of the sample was not heated. The samples were separated by SDS-PAGE, and the proteins were transferred to nitrocellulose, the serum was reacted with the strips. With the samples heated to 100°C the serum recognized a high molecular weight band of approximately 98 kDa. This is in agreement with the predicted size of Omp5, of which the pEX1-1 clone is a part, however, when the antibody was reacted to the strip with unheated EB, the pattern was different. Now a band was seen with a size of 75 kDa, in addition weaker bands were observed above the band (Figure 10). These data demonstrate that Omp5 needs boiling in SDS-sample buffer to be fully denatured and migrate with a size as predicted from the gene product. When the samples were not boiled, the protein was not fully denatured and less SDS binds to the protein and it has a more globular structure that will migrate faster in the acrylamide gel. The band pattern looked identical to what was obtained with a monoclonal antibody (MAb 26.1)(lane 6), we earlier have described (Christiansen et al., 1994), reacting with the

surface of *C. pneumoniae* EB, but the antibody do not react with the fully SDS denatured *C. pneumoniae* EB in immunoblotting.

Experimental infection of C57 black mice

Due to the realization of the altered migration of the Omp4-7 proteins without boiling, we chose to analyse antibodies against *C. pneumoniae* EBs after an experimental infection of mice. To obtain antibodies from an infection caused by *C. pneumoniae*, C57 black mice were inoculated intranasally with 10^7 CFI of *C. pneumoniae* under a light ether anaesthesia. After 14 days of infection the serum samples were obtained and the lungs were analysed for pathological changes. In two of the mice a severe pneumonia was observed in the lung sections, and in the third mouse only minor changes were observed. The serum from the mice was diluted 1:100 and reacted with purified EBs dissolved in sample buffer with and without boiling. In the preparations that had been heated to 100°C the sera from two of the mice reacted strongly with bands of 60/62 kDa and weaker bands of 55 kDa, but no reaction was observed with proteins of the size of Omp4-7 (Figure 11). However, when the sera were reacted with the preparation that had not been heated they all had a strong reaction with a broad band of an approximate size of 75 kDa. This is in agreement with the size of the Omp4-7 proteins in the unheated preparation. Therefore, it could be concluded that the epitopes of the Omp4-7 proteins recognized by the antibodies after a *C. pneumoniae* infection were discontinuous epitopes because the full denaturation of the antigen completely destroyed the epitopes. The 75 kDa protein observed in unheated samples is not Omp2 (Shown in immunoblotting with an Omp2 specific antibody)

EXAMPLE 3

Comparison of Omp4-7 of *C. pneumoniae* with putative outer membrane proteins (POMP) of *C. psittaci*

Longbottom et al. (1996) have published partial sequence from 98 to 90 kDa proteins from *C. psittaci*. They have entered the full sequence of 5 genes in this family in the EMBL database. They have named the genes "putative outer membrane proteins" (POMP) since their precise location was not determined. The family is composed of two genes that are completely identical, and two genes with high homology to these genes. They calculated a molecular size of 90 and 91 kDa. The 5th encode a protein of 98 kDa. The sequence of the Omp4-7 proteins of *C. pneumoniae* were compared to the sequences of the *C. Psittaci* POMP proteins with the programme pileup in the GCG package. The amino acid homologies were in the range of 51-63%. It is seen that the *C. pneumoniae* Omp4-5 proteins are most related to the 98 kDa POMP protein of *C. psittaci*. Interestingly, the 98 kDa *C. psittaci* POMP protein is more related to the *C. pneumoniae* genes than to the other *C. psittaci* genes. The repeated sequences of GGAI were conserved in the 98 kDa POMP protein, but only three GGAI repeats were present in the 90 and 91 kDa *C. psittaci* POMP proteins. For *C.psittaci* it has been shown that antibodies to these proteins seem to be protective for the infection.

REFERENCES

1. Caldwell,H.D., J. Kromhout and J. Schachter, Infect. Immun. 31, 1161-1176 (1981).
2. Campbell, L.A., M.P. Melgosa, D.J. Hamilton, C.-C. Kuo and J.T. Grayston, J. Clinical Microbiol., 30, 434-439 (1992).
3. Christiansen, G., and S. Birkelund. Eur. Microbiol. 1:24-29 (1992).

4. Christiansen, G., L. Østergaard, and S. Birkelund. Proceedings of the eigth International symposium on Human Infections, Eds. Orfila et al., pp 173-176, (1994).

5. Grayston, J.T., Kuo, C.-C., Campbell, L.A., and Vang, S.-P. Int. J. Syst. Bacteriol. 39, 88-90 (1989).

6. Grayston, J.T., C.-C. Kuo, S.-P. Wang and J. Altman. 1986. N. Engl. J. Med. 315, 161-168 (1986).

7. Kuo, C.C., L.A. Jackson, L.A. Campbell and J.T. Graystone. Clin. Microbiol. Rev. 8, 451-461 (1995).

8. Longbottom, D., M. Russell, G.E Jones, A. Lainson, and A.J. Herring. FEMS Microbiol. Lett. 142, 277-281 (1996).

9. Melgosa, M.P., C.-C. Kuo and L.A. Campbell, FEMS Microbiol. Lett. 112, 199-204 (1993).

10. Miyashita, N. and A. Matsumoto. J. Clin. Microbiol. 30, 2911-2916 (1992).

11. Wang, S.P., and J.T. Grayston, Am. J. Ophtalmol. 70, 367-374 (1970).

12. Freund, E.A., H. Ernø and R.M. Lemcke. Identification of mycoplasma, P377-443 in I. Norris and J.R. Bergen; Methods in Microbiology vol 13, A.P. Inc. London 1979)

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT
(A) NAME: Svend Birkelund
(B) STREET: Dept. of Medical Microbiology and Immunology, University of Århus
(C) CITY: Århus C
(D) STATE OR PROVINCE:
(E) COUNTRY: Denmark
(F) POSTAL CODE: 8000

(ii) TITLE OF THE INVENTION: Chlamydia pneumoniae antigens

(iii) NUMBER OF SEQUENCES: 8

(iv) COMPUTER-READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(v) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3200 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 205...2987
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CAATGTCGAA GAGAGCACTA ACCAGGAAAA TTGCGATTTC ATAAACCCAC TTTATTATTA    60
AATTCTTACT TGCGTCATAT AAAATAGAAA ACTCAGAGAG TCAAGATAAA AATTCTTGAC   120
AGCTGTTTTG TCATCTTTAA CTTGATTTAC TTATTTTGTT TCTATATTGA TGCGAATAGT   180
TCTCTAAAAA ACAAAAGCAT TACC ATG AAG ACT TCG ATT CCT TGG GTT TTA       231
                           Met Lys Thr Ser Ile Pro Trp Val Leu
                            1               5

GTT TCC TCC GTG TTA GCT TTC TCA TGT CAC CTA CAG TCA CTA GCT AAC      279
Val Ser Ser Val Leu Ala Phe Ser Cys His Leu Gln Ser Leu Ala Asn
10                  15                  20                  25

GAG GAA CTT TTA TCA CCT GAT GAT AGC TTT AAT GGA AAT ATC GAT TCA      327
Glu Glu Leu Leu Ser Pro Asp Asp Ser Phe Asn Gly Asn Ile Asp Ser
                30                  35                  40
```

```
GGA ACG TTT ACT CCA AAA ACT TCA GCC ACA ACA TAT TCT CTA ACA GGA      375
Gly Thr Phe Thr Pro Lys Thr Ser Ala Thr Thr Tyr Ser Leu Thr Gly
            45                  50                  55

GAT GTC TTC TTT TAC GAG CCT GGA AAA GGC ACT CCC TTA TCT GAC AGT      423
Asp Val Phe Phe Tyr Glu Pro Gly Lys Gly Thr Pro Leu S r Asp Ser
        60                  65                  70

TGT TTT AAG CAA ACC ACG GAC AAT CTT ACC TTC TTG GGG AAC GGT CAT      471
Cys Phe Lys Gln Thr Thr Asp Asn Leu Thr Phe Leu Gly Asn Gly His
    75                  80                  85

AGC TTA ACG TTT GGC TTT ATA GAT GCT GGC ACT CAT GCA GGT GCT GCT      519
Ser Leu Thr Phe Gly Phe Ile Asp Ala Gly Thr His Ala Gly Ala Ala
90                  95                  100                 105

GCA TCT ACA ACA GCA AAT AAG AAT CTT ACC TTC TCA GGG TTT TCC TTA      567
Ala Ser Thr Thr Ala Asn Lys Asn Leu Thr Phe Ser Gly Phe Ser Leu
                110                 115                 120

CTG AGT TTT GAT TCC TCT CCT AGC ACA ACG GTT ACT ACA GGT CAG GGA      615
Leu Ser Phe Asp Ser Ser Pro Ser Thr Thr Val Thr Thr Gly Gln Gly
            125                 130                 135

ACG CTT TCC TCA GCA GGA GGC GTA AAT TTA GAA AAT ATT CGT AAA CTT      663
Thr Leu Ser Ser Ala Gly Gly Val Asn Leu Glu Asn Ile Arg Lys Leu
        140                 145                 150

GTA GTT GCT GGG AAT TTT TCT ACT GCA GAT GGT GGA GCT ATC AAA GGA      711
Val Val Ala Gly Asn Phe Ser Thr Ala Asp Gly Gly Ala Ile Lys Gly
    155                 160                 165

GCG TCT TTC CTT TTA ACT GGC ACT TCT GGA GAT GCT CTT TTT AGT AAC      759
Ala Ser Phe Leu Leu Thr Gly Thr Ser Gly Asp Ala Leu Phe Ser Asn
170                 175                 180                 185

AAC TCT TCA TCA ACA AAG GGA GGA GCA ATT GCT ACT ACA GCA GGC GCT      807
Asn Ser Ser Ser Thr Lys Gly Gly Ala Ile Ala Thr Thr Ala Gly Ala
                190                 195                 200

CGC ATA GCA AAT AAC ACA GGT TAT GTT AGA TTC CTA TCT AAC ATA GCG      855
Arg Ile Ala Asn Asn Thr Gly Tyr Val Arg Phe Leu Ser Asn Ile Ala
            205                 210                 215

TCT ACG TCA GGA GGC GCT ATC GAT GAT GAA GGC ACG TCG ATA CTA TCG      903
Ser Thr Ser Gly Gly Ala Ile Asp Asp Glu Gly Thr Ser Ile Leu Ser
        220                 225                 230

AAC AAC AAA TTT CTA TAT TTT GAA GGG AAT GCA GCG AAA ACT ACT GGC      951
Asn Asn Lys Phe Leu Tyr Phe Glu Gly Asn Ala Ala Lys Thr Thr Gly
    235                 240                 245

GGT GCG ATC TGC AAC ACC AAG GCG AGT GGA TCT CCT GAA CTG ATA ATC      999
Gly Ala Ile Cys Asn Thr Lys Ala Ser Gly Ser Pro Glu Leu Ile Ile
250                 255                 260                 265
```

```
TCT AAC AAT AAG ACT CTG ATC TTT GCT TCA AAC GTA GCA GAA ACA AGC     1047
Ser Asn Asn Lys Thr Leu Ile Phe Ala Ser Asn Val Ala Glu Thr Ser
                270                 275                 280

GGT GGC GCC ATC CAT GCT AAA AAG CTA GCC CTT TCC TCT GGA GGC TTT     1095
Gly Gly Ala Ile His Ala Lys Lys Leu Ala Leu Ser Ser Gly Gly Phe
            285                 290                 295

ACA GAG TTT CTA CGA AAT AAT GTC TCA TCA GCA ACT CCT AAG GGG GGT     1143
Thr Glu Phe Leu Arg Asn Asn Val Ser Ser Ala Thr Pro Lys Gly Gly
        300                 305                 310

GCT ATC AGC ATC GAT GCC TCA GGA GAG CTC AGT CTT TCT GCA GAG ACA     1191
Ala Ile Ser Ile Asp Ala Ser Gly Glu Leu Ser Leu Ser Ala Glu Thr
    315                 320                 325

GGA AAC ATT ACC TTT GTA AGA AAT ACC CTT ACA ACA ACC GGA AGT ACC     1239
Gly Asn Ile Thr Phe Val Arg Asn Thr Leu Thr Thr Thr Gly Ser Thr
330                 335                 340                 345

GAT ACT CCT AAA CGT AAT GCG ATC AAC ATA GGA AGT AAC GGG AAA TTC     1287
Asp Thr Pro Lys Arg Asn Ala Ile Asn Ile Gly Ser Asn Gly Lys Phe
                350                 355                 360

ACG GAA TTA CGG GCT GCT AAA AAT CAT ACA ATT TTC TTC TAT GAT CCC     1335
Thr Glu Leu Arg Ala Ala Lys Asn His Thr Ile Phe Phe Tyr Asp Pro
            365                 370                 375

ATC ACT TCA GAA GGA ACC TCA TCA GAC GTA TTG AAG ATA AAT AAC GGC     1383
Ile Thr Ser Glu Gly Thr Ser Ser Asp Val Leu Lys Ile Asn Asn Gly
        380                 385                 390

TCT GCG GGA GCT CTC AAT CCA TAT CAA GGA ACG ATT CTA TTT TCT GGA     1431
Ser Ala Gly Ala Leu Asn Pro Tyr Gln Gly Thr Ile Leu Phe Ser Gly
    395                 400                 405

GAA ACC CTA ACA GCA GAT GAA CTT AAA GTT GCT GAC AAT TTA AAA TCT     1479
Glu Thr Leu Thr Ala Asp Glu Leu Lys Val Ala Asp Asn Leu Lys Ser
410                 415                 420                 425

TCA TTC ACG CAG CCA GTC TCC CTA TCC GGA GGA AAG TTA TTG CTA CAA     1527
Ser Phe Thr Gln Pro Val Ser Leu Ser Gly Gly Lys Leu Leu Leu Gln
                430                 435                 440

AAG GGA GTC ACT TTA GAG AGC ACG AGC TTC TCT CAA GAG GCC GGT TCT     1575
Lys Gly Val Thr Leu Glu Ser Thr Ser Phe Ser Gln Glu Ala Gly Ser
            445                 450                 455

CTC CTC GGC ATG GAT TCA GGA ACG ACA TTA TCA ACT ACA GCT GGG AGT     1623
Leu Leu Gly Met Asp Ser Gly Thr Thr Leu Ser Thr Thr Ala Gly Ser
        460                 465                 470

ATT ACA ATC ACG AAC CTA GGA ATC AAT GTT GAC TCC TTA GGT CTT AAG     1671
Ile Thr Ile Thr Asn Leu Gly Ile Asn Val Asp Ser Leu Gly Leu Lys
    475                 480                 485
```

```
CAG CCC GTC AGC CTA ACA GCA AAA GGT GCT TCA AAT AAA GTG ATC GTA    1719
Gln Pro Val Ser Leu Thr Ala Lys Gly Ala Ser Asn Lys Val Ile Val
490                 495                 500                 505

TCT GGG AAG CTC AAC CTG ATT GAT ATT GAA GGG AAC ATT TAT GAA AGT    1767
Ser Gly Lys Leu Asn Leu Ile Asp Ile Glu Gly Asn Ile Tyr Glu Ser
                510                 515                 520

CAT ATG TTC AGC CAT GAC CAG CTC TTC TCT CTA TTA AAA ATC ACG GTT    1815
His Met Phe Ser His Asp Gln Leu Phe Ser Leu Leu Lys Ile Thr Val
            525                 530                 535

GAT GCT GAT GTT GAT ACT AAC GTT GAC ATC AGC AGC CTT ATC CCT GTT    1863
Asp Ala Asp Val Asp Thr Asn Val Asp Ile Ser Ser Leu Ile Pro Val
        540                 545                 550

CCT GCT GAG GAT CCT AAT TCA GAA TAC GGA TTC CAA GGA CAA TGG AAT    1911
Pro Ala Glu Asp Pro Asn Ser Glu Tyr Gly Phe Gln Gly Gln Trp Asn
    555                 560                 565

GTT AAT TGG ACT ACG GAT ACA GCT ACA AAT ACA AAA GAG GCC ACG GCA    1959
Val Asn Trp Thr Thr Asp Thr Ala Thr Asn Thr Lys Glu Ala Thr Ala
570                 575                 580                 585

ACT TGG ACC AAA ACA GGA TTT GTT CCC AGC CCC GAA AGA AAA TCT GCG    2007
Thr Trp Thr Lys Thr Gly Phe Val Pro Ser Pro Glu Arg Lys Ser Ala
                590                 595                 600

TTA GTA TGC AAT ACC CTA TGG GGA GTC TTT ACT GAC ATT CGC TCT CTG    2055
Leu Val Cys Asn Thr Leu Trp Gly Val Phe Thr Asp Ile Arg Ser Leu
            605                 610                 615

CAA CAG CTT GTA GAG ATC GGC GCA ACT GGT ATG GAA CAC AAA CAA GGT    2103
Gln Gln Leu Val Glu Ile Gly Ala Thr Gly Met Glu His Lys Gln Gly
        620                 625                 630

TTC TGG GTT TCC TCC ATG ACG AAC TTC CTG CAT AAG ACT GGA GAT GAA    2151
Phe Trp Val Ser Ser Met Thr Asn Phe Leu His Lys Thr Gly Asp Glu
    635                 640                 645

AAT CGC AAA GGC TTC CGT CAT ACC TCT GGA GGC TAC GTC ATC GGT GGA    2199
Asn Arg Lys Gly Phe Arg His Thr Ser Gly Gly Tyr Val Ile Gly Gly
650                 655                 660                 665

AGT GCT CAC ACT CCT AAA GAC GAC CTA TTT ACC TTT GCG TTC TGC CAT    2247
Ser Ala His Thr Pro Lys Asp Asp Leu Phe Thr Phe Ala Phe Cys His
                670                 675                 680

CTC TTT GCT AGA GAC AAA GAT TGT TTT ATC GCT CAC AAC AAC TCT AGA    2295
Leu Phe Ala Arg Asp Lys Asp Cys Phe Ile Ala His Asn Asn Ser Arg
            685                 690                 695

ACC TAC GGT GGA ACT TTA TTC TTC AAG CAC TCT CAT ACC CTA CAA CCC    2343
Thr Tyr Gly Gly Thr Leu Phe Phe Lys His Ser His Thr Leu Gln Pro
        700                 705                 710
```

```
CAA AAC TAT TTG AGA TTA GGA AGA GCA AAG TTT TCT GAA TCA GCT ATA    2391
Gln Asn Tyr Leu Arg Leu Gly Arg Ala Lys Phe Ser Glu Ser Ala Ile
    715                 720                 725

GAA AAA TTC CCT AGG GAA ATT CCC CTA GCC TTG GAT GTC CAA GTT TCG    2439
Glu Lys Phe Pro Arg Glu Ile Pro Leu Ala Leu Asp Val Gln Val Ser
730                 735                 740                 745

TTC AGC CAT TCA GAC AAC CGT ATG GAA ACG CAC TAT ACC TCA TTG CCA    2487
Phe Ser His Ser Asp Asn Arg Met Glu Thr His Tyr Thr Ser Leu Pro
                750                 755                 760

GAA TCC GAA GGT TCT TGG AGC AAC GAG TGT ATA GCT GGT GGT ATC GGC    2535
Glu Ser Glu Gly Ser Trp Ser Asn Glu Cys Ile Ala Gly Gly Ile Gly
            765                 770                 775

CTA GAC CTT CCT TTT GTT CTT TCC AAC CCA CAT CCT CTT TTC AAG ACC    2583
Leu Asp Leu Pro Phe Val Leu Ser Asn Pro His Pro Leu Phe Lys Thr
        780                 785                 790

TTC ATT CCA CAG ATG AAA GTC GAA ATG GTT TAT GTA TCA CAA AAT AGC    2631
Phe Ile Pro Gln Met Lys Val Glu Met Val Tyr Val Ser Gln Asn Ser
    795                 800                 805

TTC TTC GAA AGC TCT AGT GAT GGC CGT GGT TTT AGT ATT GGA AGG CTG    2679
Phe Phe Glu Ser Ser Ser Asp Gly Arg Gly Phe Ser Ile Gly Arg Leu
810                 815                 820                 825

CTT AAC CTC TCG ATT CCT GTG GGT GCG AAA TTC GTG CAG GGG GAT ATC    2727
Leu Asn Leu Ser Ile Pro Val Gly Ala Lys Phe Val Gln Gly Asp Ile
                830                 835                 840

GGA GAT TCC TAC ACC TAT GAT CTC TCA GGA TTC TTT GTT TCC GAT GTC    2775
Gly Asp Ser Tyr Thr Tyr Asp Leu Ser Gly Phe Phe Val Ser Asp Val
            845                 850                 855

TAT CGT AAC AAT CCC CAA TCT ACA GCG ACT CTT GTG ATG AGC CCA GAC    2823
Tyr Arg Asn Asn Pro Gln Ser Thr Ala Thr Leu Val Met Ser Pro Asp
        860                 865                 870

TCT TGG AAA ATT CGC GGT GGC AAT CTT TCA AGA CAG GCA TTT TTA CTG    2871
Ser Trp Lys Ile Arg Gly Gly Asn Leu Ser Arg Gln Ala Phe Leu Leu
    875                 880                 885

AGG GGT AGC AAC AAC TAC GTC TAC AAC TCC AAT TGT GAG CTC TTC GGA    2919
Arg Gly Ser Asn Asn Tyr Val Tyr Asn Ser Asn Cys Glu Leu Phe Gly
890                 895                 900                 905

CAT TAC GCT ATG GAA CTC CGT GGA TCT TCA AGG AAC TAC AAT GTA GAT    2967
His Tyr Ala Met Glu Leu Arg Gly Ser Ser Arg Asn Tyr Asn Val Asp
                910                 915                 920

GTT GGT ACC AAA CTC CGA TT CTAGATTGCT AAAACTCCCT AGTTCTTCTA GGGAG  3022
Val Gly Thr Lys Leu Arg Phe
            925

TTTTCTCATA CTTTTAGGGA AATATTTGCT ATAGGGAATG CTTTCCTTGC AAACTGTAAA  3082
AAATAACATT TGTCCCTCTT CAAAAAAGAT TTCTTTTAAT AATTTCTAGT TATAATTTTA  3142
```

TTTTAAAAAC AGTTAAATAA TTAATAGACA ATAATCTATT CTTATTGACT TCTTTTTT 3200

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 928 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Thr Ser Ile Pro Trp Val Leu Val Ser Ser Val Leu Ala Phe
 1               5                  10                  15
Ser Cys His Leu Gln Ser Leu Ala Asn Glu Glu Leu Leu Ser Pro Asp
            20                  25                  30
Asp Ser Phe Asn Gly Asn Ile Asp Ser Gly Thr Phe Thr Pro Lys Thr
        35                  40                  45
Ser Ala Thr Thr Tyr Ser Leu Thr Gly Asp Val Phe Phe Tyr Glu Pro
    50                  55                  60
Gly Lys Gly Thr Pro Leu Ser Asp Ser Cys Phe Lys Gln Thr Thr Asp
65                  70                  75                  80
Asn Leu Thr Phe Leu Gly Asn Gly His Ser Leu Thr Phe Gly Phe Ile
                85                  90                  95
Asp Ala Gly Thr His Ala Gly Ala Ala Ala Ser Thr Thr Ala Asn Lys
            100                 105                 110
Asn Leu Thr Phe Ser Gly Phe Ser Leu Leu Ser Phe Asp Ser Ser Pro
        115                 120                 125
Ser Thr Thr Val Thr Thr Gly Gln Gly Thr Leu Ser Ser Ala Gly Gly
    130                 135                 140
Val Asn Leu Glu Asn Ile Arg Lys Leu Val Val Ala Gly Asn Phe Ser
145                 150                 155                 160
Thr Ala Asp Gly Gly Ala Ile Lys Gly Ala Ser Phe Leu Leu Thr Gly
                165                 170                 175
Thr Ser Gly Asp Ala Leu Phe Ser Asn Asn Ser Ser Ser Thr Lys Gly
            180                 185                 190
Gly Ala Ile Ala Thr Thr Ala Gly Ala Arg Ile Ala Asn Asn Thr Gly
        195                 200                 205
Tyr Val Arg Phe Leu Ser Asn Ile Ala Ser Thr Ser Gly Gly Ala Ile
    210                 215                 220
Asp Asp Glu Gly Thr Ser Ile Leu Ser Asn Asn Lys Phe Leu Tyr Phe
225                 230                 235                 240
Glu Gly Asn Ala Ala Lys Thr Thr Gly Gly Ala Ile Cys Asn Thr Lys
                245                 250                 255
Ala Ser Gly Ser Pro Glu Leu Ile Ile Ser Asn Asn Lys Thr Leu Ile
            260                 265                 270
Phe Ala Ser Asn Val Ala Glu Thr Ser Gly Gly Ala Ile His Ala Lys
        275                 280                 285
Lys Leu Ala Leu Ser Ser Gly Gly Phe Thr Glu Phe Leu Arg Asn Asn
    290                 295                 300
Val Ser Ser Ala Thr Pro Lys Gly Gly Ala Ile Ser Ile Asp Ala Ser
305                 310                 315                 320
Gly Glu Leu Ser Leu Ser Ala Glu Thr Gly Asn Ile Thr Phe Val Arg
                325                 330                 335
```

```
Asn Thr Leu Thr Thr Thr Gly Ser Thr Asp Thr Pro Lys Arg Asn Ala
            340                 345                 350
Ile Asn Ile Gly Ser Asn Gly Lys Phe Thr Glu Leu Arg Ala Ala Lys
        355                 360                 365
Asn His Thr Ile Phe Phe Tyr Asp Pro Ile Thr Ser Glu Gly Thr Ser
    370                 375                 380
Ser Asp Val Leu Lys Ile Asn Asn Gly Ser Ala Gly Ala Leu Asn Pro
385                 390                 395                 400
Tyr Gln Gly Thr Ile Leu Phe Ser Gly Glu Thr Leu Thr Ala Asp Glu
                405                 410                 415
Leu Lys Val Ala Asp Asn Leu Lys Ser Ser Phe Thr Gln Pro Val Ser
            420                 425                 430
Leu Ser Gly Gly Lys Leu Leu Leu Gln Lys Gly Val Thr Leu Glu Ser
        435                 440                 445
Thr Ser Phe Ser Gln Glu Ala Gly Ser Leu Leu Gly Met Asp Ser Gly
    450                 455                 460
Thr Thr Leu Ser Thr Thr Ala Gly Ser Ile Thr Ile Thr Asn Leu Gly
465                 470                 475                 480
Ile Asn Val Asp Ser Leu Gly Leu Lys Gln Pro Val Ser Leu Thr Ala
                485                 490                 495
Lys Gly Ala Ser Asn Lys Val Ile Val Ser Gly Lys Leu Asn Leu Ile
            500                 505                 510
Asp Ile Glu Gly Asn Ile Tyr Glu Ser His Met Phe Ser His Asp Gln
        515                 520                 525
Leu Phe Ser Leu Leu Lys Ile Thr Val Asp Ala Asp Val Asp Thr Asn
    530                 535                 540
Val Asp Ile Ser Ser Leu Ile Pro Val Pro Ala Glu Asp Pro Asn Ser
545                 550                 555                 560
Glu Tyr Gly Phe Gln Gly Gln Trp Asn Val Asn Trp Thr Thr Asp Thr
                565                 570                 575
Ala Thr Asn Thr Lys Glu Ala Thr Ala Thr Trp Thr Lys Thr Gly Phe
            580                 585                 590
Val Pro Ser Pro Glu Arg Lys Ser Ala Leu Val Cys Asn Thr Leu Trp
        595                 600                 605
Gly Val Phe Thr Asp Ile Arg Ser Leu Gln Gln Leu Val Glu Ile Gly
    610                 615                 620
Ala Thr Gly Met Glu His Lys Gln Gly Phe Trp Val Ser Ser Met Thr
625                 630                 635                 640
Asn Phe Leu His Lys Thr Gly Asp Glu Asn Arg Lys Gly Phe Arg His
                645                 650                 655
Thr Ser Gly Gly Tyr Val Ile Gly Gly Ser Ala His Thr Pro Lys Asp
            660                 665                 670
Asp Leu Phe Thr Phe Ala Phe Cys His Leu Phe Ala Arg Asp Lys Asp
        675                 680                 685
Cys Phe Ile Ala His Asn Asn Ser Arg Thr Tyr Gly Gly Thr Leu Phe
    690                 695                 700
Phe Lys His Ser His Thr Leu Gln Pro Gln Asn Tyr Leu Arg Leu Gly
705                 710                 715                 720
Arg Ala Lys Phe Ser Glu Ser Ala Ile Glu Lys Phe Pro Arg Glu Ile
                725                 730                 735
Pro Leu Ala Leu Asp Val Gln Val Ser Phe Ser His Ser Asp Asn Arg
            740                 745                 750
Met Glu Thr His Tyr Thr Ser Leu Pro Glu Ser Glu Gly Ser Trp Ser
        755                 760                 765
Asn Glu Cys Ile Ala Gly Gly Ile Gly Leu Asp Leu Pro Phe Val Leu
    770                 775                 780
Ser Asn Pro His Pro Leu Phe Lys Thr Phe Ile Pro Gln Met Lys Val
785                 790                 795                 800
```

```
Glu Met Val Tyr Val Ser Gln Asn Ser Phe Phe Glu Ser Ser Ser Asp
                805                 810                 815
Gly Arg Gly Phe Ser Ile Gly Arg Leu Leu Asn Leu Ser Ile Pro Val
            820                 825                 830
Gly Ala Lys Phe Val Gln Gly Asp Ile Gly Asp Ser Tyr Thr Tyr Asp
        835                 840                 845
Leu Ser Gly Phe Phe Val Ser Asp Val Tyr Arg Asn Asn Pro Gln Ser
    850                 855                 860
Thr Ala Thr Leu Val Met Ser Pro Asp Ser Trp Lys Ile Arg Gly Gly
865                 870                 875                 880
Asn Leu Ser Arg Gln Ala Phe Leu Leu Arg Gly Ser Asn Asn Tyr Val
                885                 890                 895
Tyr Asn Ser Asn Cys Glu Leu Phe Gly His Tyr Ala Met Glu Leu Arg
            900                 905                 910
Gly Ser Ser Arg Asn Tyr Asn Val Asp Val Gly Thr Lys Leu Arg Phe
        915                 920                 925
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3000 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: Coding Sequence
- (B) LOCATION: 259...3000
- (D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATCAGGTGAT AAAAGTTCCT CGTTAGCTAG TGACTGTAGG TGACATGAGA AAGCTAACAC     60
GGAGGAAACT AAAACCCAAG GAATCGAAGT CTTCATGGTA ATGCTTTTGT TTTTTAGAGA    120
ACTATTCGCA TCAATATAGA AACAAAATAA GTAAATCAAG TTAAAGATGA CAAAACAGCT    180
GTCAAGAATT TTTATCTTGA CTCTCTGAGT TTTCTATTTT ATATGACGCA AGTAAGAATT    240
TAATAATAAA GTGGGTTT ATG AAA TCG CAA TTT TCC TGG TTA GTG CTC TCT      291
                    Met Lys Ser Gln Phe Ser Trp Leu Val Leu Ser
                     1               5                  10

TCG ACA TTG GCA TGT TTT ACT AGT TGT TCC ACT GTT TTT GCT GCA ACT      339
Ser Thr Leu Ala Cys Phe Thr Ser Cys Ser Thr Val Phe Ala Ala Thr
            15                  20                  25

GCT GAA AAT ATA GGC CCC TCT GAT AGC TTT GAC GGA AGT ACT AAC ACA      387
Ala Glu Asn Ile Gly Pro Ser Asp Ser Phe Asp Gly Ser Thr Asn Thr
        30                  35                  40

GGC ACC TAT ACT CCT AAA AAT ACG ACT ACT GGA ATA GAC TAT ACT CTG      435
Gly Thr Tyr Thr Pro Lys Asn Thr Thr Thr Gly Ile Asp Tyr Thr Leu
    45                  50                  55

ACA GGA GAT ATA ACT CTG CAA AAC CTT GGG GAT TCG GCA GCT TTA ACG      483
Thr Gly Asp Ile Thr Leu Gln Asn Leu Gly Asp Ser Ala Ala Leu Thr
60                  65                  70                  75
```

```
AAG GGT TGT TTT TCT GAC ACT ACG GAA TCT TTA AGC TTT GCC GGT AAG      531
Lys Gly Cys Phe Ser Asp Thr Thr Glu Ser Leu Ser Phe Ala Gly Lys
                80                  85                  90

GGG TAC TCA CTT TCT TTT TTA AAT ATT AAG TCT AGT GCT GAA GGC GCA      579
Gly Tyr Ser Leu Ser Phe Leu Asn Ile Lys Ser Ser Ala Glu Gly Ala
            95                  100                 105

GCA CTT TCT GTT ACA ACT GAT AAA AAT CTG TCG CTA ACA GGA TTT TCG      627
Ala Leu Ser Val Thr Thr Asp Lys Asn Leu Ser Leu Thr Gly Phe Ser
        110                 115                 120

AGT CTT ACT TTC TTA GCG GCC CCA TCA TCG GTA ATC ACA ACC CCC TCA      675
Ser Leu Thr Phe Leu Ala Ala Pro Ser Ser Val Ile Thr Thr Pro Ser
    125                 130                 135

GGA AAA GGT GCA GTT AAA TGT GGA GGG GAT CTT ACA TTT GAT AAC AAT      723
Gly Lys Gly Ala Val Lys Cys Gly Gly Asp Leu Thr Phe Asp Asn Asn
140                 145                 150                 155

GGA ACT ATT TTA TTT AAA CAA GAT TAC TGT GAG GAA AAT GGC GGA GCC      771
Gly Thr Ile Leu Phe Lys Gln Asp Tyr Cys Glu Glu Asn Gly Gly Ala
                160                 165                 170

ATT TCT ACC AAG AAT CTT TCT TTG AAA AAC AGC ACG GGA TCG ATT TCT      819
Ile Ser Thr Lys Asn Leu Ser Leu Lys Asn Ser Thr Gly Ser Ile Ser
            175                 180                 185

TTT GAA GGG AAT AAA TCG AGC GCA ACA GGG AAA AAA GGT GGG GCT ATT      867
Phe Glu Gly Asn Lys Ser Ser Ala Thr Gly Lys Lys Gly Gly Ala Ile
        190                 195                 200

TGT GCT ACT GGT ACT GTA GAT ATT ACA AAT AAT ACG GCT CCT ACC CTC      915
Cys Ala Thr Gly Thr Val Asp Ile Thr Asn Asn Thr Ala Pro Thr Leu
    205                 210                 215

TTC TCG AAC AAT ATT GCT GAA GCT GCA GGT GGA GCT ATA AAT AGC ACA      963
Phe Ser Asn Asn Ile Ala Glu Ala Ala Gly Gly Ala Ile Asn Ser Thr
220                 225                 230                 235

GGA AAC TGT ACA ATT ACA GGG AAT ACG TCT CTT GTA TTT TCT GAA AAT     1011
Gly Asn Cys Thr Ile Thr Gly Asn Thr Ser Leu Val Phe Ser Glu Asn
                240                 245                 250

AGT GTG ACA GCG ACC GCA GGA AAT GGA GGA GCT CTT TCT GGA GAT GCC     1059
Ser Val Thr Ala Thr Ala Gly Asn Gly Gly Ala Leu Ser Gly Asp Ala
            255                 260                 265

GAT GTT ACC ATA TCT GGG AAT CAG AGT GTA ACT TTC TCA GGA AAC CAA     1107
Asp Val Thr Ile Ser Gly Asn Gln Ser Val Thr Phe Ser Gly Asn Gln
        270                 275                 280

GCT GTA GCT AAT GGC GGA GCC ATT TAT GCT AAG AAG CTT ACA CTG GCT     1155
Ala Val Ala Asn Gly Gly Ala Ile Tyr Ala Lys Lys Leu Thr Leu Ala
    285                 290                 295
```

```
TCC GGG GGG GGG GGG GGT ATC TCC TTT TCT AAC AAT ATA GTC CAA GGT    1203
Ser Gly Gly Gly Gly Gly Ile Ser Phe Ser Asn Asn Ile Val Gln Gly
300                 305                 310                 315

ACC ACT GCA GGT AAT GGT GGA GCC ATT TCT ATA CTG GCA GCT GGA GAG    1251
Thr Thr Ala Gly Asn Gly Gly Ala Ile Ser Ile Leu Ala Ala Gly Glu
                320                 325                 330

TGT AGT CTT TCA GCA GAA GCA GGG GAC ATT ACC TTC AAT GGG AAT GCC    1299
Cys Ser Leu Ser Ala Glu Ala Gly Asp Ile Thr Phe Asn Gly Asn Ala
            335                 340                 345

ATT GTT GCA ACT ACA CCA CAA ACT ACA AAA AGA AAT TCT ATT GAC ATA    1347
Ile Val Ala Thr Thr Pro Gln Thr Thr Lys Arg Asn Ser Ile Asp Ile
        350                 355                 360

GGA TCT ACT GCA AAG ATC ACG AAT TTA CGT GCA ATA TCT GGG CAT AGC    1395
Gly Ser Thr Ala Lys Ile Thr Asn Leu Arg Ala Ile Ser Gly His Ser
    365                 370                 375

ATC TTT TTC TAC GAT CCG ATT ACT GCT AAT ACG GCT GCG GAT TCT ACA    1443
Ile Phe Phe Tyr Asp Pro Ile Thr Ala Asn Thr Ala Ala Asp Ser Thr
380                 385                 390                 395

GAT ACT TTA AAT CTC AAT AAG GCT GAT GCA GGT AAT AGT ACA GAT TAT    1491
Asp Thr Leu Asn Leu Asn Lys Ala Asp Ala Gly Asn Ser Thr Asp Tyr
                400                 405                 410

AGT GGG TCG ATT GTT TTT TCT GGT GAA AAG CTC TCT GAA GAT GAA GCA    1539
Ser Gly Ser Ile Val Phe Ser Gly Glu Lys Leu Ser Glu Asp Glu Ala
            415                 420                 425

AAA GTT GCA GAC AAC CTC ACT TCT ACG CTG AAG CAG CCT GTA ACT CTA    1587
Lys Val Ala Asp Asn Leu Thr Ser Thr Leu Lys Gln Pro Val Thr Leu
        430                 435                 440

ACT GCA GGA AAT TTA GTA CTT AAA CGT GGT GTC ACT CTC GAT ACG AAA    1635
Thr Ala Gly Asn Leu Val Leu Lys Arg Gly Val Thr Leu Asp Thr Lys
    445                 450                 455

GGC TTT ACT CAG ACC GCG GGT TCC TCT GTT ATT ATG GAT GCG GGC ACA    1683
Gly Phe Thr Gln Thr Ala Gly Ser Ser Val Ile Met Asp Ala Gly Thr
460                 465                 470                 475

ACG TTA AAA GCA AGT ACA GAG GAG GTC ACT TTA ACA GGT CTT TCC ATT    1731
Thr Leu Lys Ala Ser Thr Glu Glu Val Thr Leu Thr Gly Leu Ser Ile
                480                 485                 490

CCT GTA GAC TCT TTA GGC GAG GGT AAG AAA GTT GTA ATT GCT GCT TCT    1779
Pro Val Asp Ser Leu Gly Glu Gly Lys Lys Val Val Ile Ala Ala Ser
            495                 500                 505

GCA GCA AGT AAA AAT GTA GCC CTT AGT GGT CCG ATT CTT CTT TTG GAT    1827
Ala Ala Ser Lys Asn Val Ala Leu Ser Gly Pro Ile Leu Leu Leu Asp
        510                 515                 520
```

```
AAC CAA GGG AAT GCT TAT GAA AAT CAC GAC TTA GGA AAA ACT CAA GAC     1875
Asn Gln Gly Asn Ala Tyr Glu Asn His Asp Leu Gly Lys Thr Gln Asp
    525                 530                 535

TTT TCA TTT GTG CAG CTC TCT GCT CTG GGT ACT GCA ACA ACT ACA GAT     1923
Phe Ser Phe Val Gln Leu Ser Ala Leu Gly Thr Ala Thr Thr Thr Asp
540                 545                 550                 555

GTT CCA GCG GTT CCT ACA GTA GCA ACT CCT ACG CAC TAT GGG TAT CAA     1971
Val Pro Ala Val Pro Thr Val Ala Thr Pro Thr His Tyr Gly Tyr Gln
                560                 565                 570

GGT ACT TGG GGA ATG ACT TGG GTT GAT GAT ACC GCA AGC ACT CCA AAG     2019
Gly Thr Trp Gly Met Thr Trp Val Asp Asp Thr Ala Ser Thr Pro Lys
            575                 580                 585

ACT AAG ACA GCG ACA TTA GCT TGG ACC AAT ACA GGC TAC CTT CCG AAT     2067
Thr Lys Thr Ala Thr Leu Ala Trp Thr Asn Thr Gly Tyr Leu Pro Asn
        590                 595                 600

CCT GAG CGT CAA GGA CCT TTA GTT CCT AAT AGC CTT TGG GGA TCT TTT     2115
Pro Glu Arg Gln Gly Pro Leu Val Pro Asn Ser Leu Trp Gly Ser Phe
    605                 610                 615

TCA GAC ATC CAA GCG ATT CAA GGT GTC ATA GAG AGA AGT GCT TTG ACT     2163
Ser Asp Ile Gln Ala Ile Gln Gly Val Ile Glu Arg Ser Ala Leu Thr
620                 625                 630                 635

CTT TGT TCA GAT CGA GGC TTC TGG GCT GCG GGA GTC GCC AAT TTC TTA     2211
Leu Cys Ser Asp Arg Gly Phe Trp Ala Ala Gly Val Ala Asn Phe Leu
                640                 645                 650

GAT AAA GAT AAG AAA GGG GAA AAA CGC AAA TAC CGT CAT AAA TCT GGT     2259
Asp Lys Asp Lys Lys Gly Glu Lys Arg Lys Tyr Arg His Lys Ser Gly
            655                 660                 665

GGA TAT GCT ATC GGA GGT GCA GCG CAA ACT TGT TCT GAA AAC TTA ATT     2307
Gly Tyr Ala Ile Gly Gly Ala Ala Gln Thr Cys Ser Glu Asn Leu Ile
        670                 675                 680

AGC TTT GCC TTT TGC CAA CTC TTT GGT AGC GAT AAA GAT TTC TTA GTC     2355
Ser Phe Ala Phe Cys Gln Leu Phe Gly Ser Asp Lys Asp Phe Leu Val
    685                 690                 695

GCT AAA AAT CAT ACT GAT ACC TAT GCA GGA GCC TTC TAT ATC CAA CAC     2403
Ala Lys Asn His Thr Asp Thr Tyr Ala Gly Ala Phe Tyr Ile Gln His
700                 705                 710                 715

ATT ACA GAA TGT AGT GGG TTC ATA GGT TGT CTC TTA GAT AAA CTT CCT     2451
Ile Thr Glu Cys Ser Gly Phe Ile Gly Cys Leu Leu Asp Lys Leu Pro
                720                 725                 730

GGC TCT TGG AGT CAT AAA CCC CTC GTT TTA GAA GGG CAG CTC GCT TAT     2499
Gly Ser Trp Ser His Lys Pro Leu Val Leu Glu Gly Gln Leu Ala Tyr
            735                 740                 745
```

```
AGC CAC GTC AGT AAT GAT CTG AAG ACA AAG TAT ACT GCG TAT CCT GAG     2547
Ser His Val Ser Asn Asp Leu Lys Thr Lys Tyr Thr Ala Tyr Pro Glu
        750                 755                 760

GTG AAA GGT TCT TGG GGG AAT AAT GCT TTT AAC ATG ATG TTG GGA GCT     2595
Val Lys Gly Ser Trp Gly Asn Asn Ala Phe Asn Met Met Leu Gly Ala
    765                 770                 775

TCT TCT CAT TCT TAT CCT GAA TAC CTG CAT TGT TTT GAT ACC TAT GCT     2643
Ser Ser His Ser Tyr Pro Glu Tyr Leu His Cys Phe Asp Thr Tyr Ala
780                 785                 790                 795

CCA TAC ATC AAA CTG AAT CTG ACC TAT ATA CGT CAG GAC AGC TTC TCG     2691
Pro Tyr Ile Lys Leu Asn Leu Thr Tyr Ile Arg Gln Asp Ser Phe Ser
                800                 805                 810

GAG AAA GGT ACA GAA GGA AGA TCT TTT GAT GAC AGC AAC CTC TTC AAT     2739
Glu Lys Gly Thr Glu Gly Arg Ser Phe Asp Asp Ser Asn Leu Phe Asn
            815                 820                 825

TTA TCT TTG CCT ATA GGG GTG AAG TTT GAG AAG TTC TCT GAT TGT AAT     2787
Leu Ser Leu Pro Ile Gly Val Lys Phe Glu Lys Phe Ser Asp Cys Asn
        830                 835                 840

GAC TTT TCT TAT GAT CTG ACT TTA TCC TAT GTT CCT GAT CTT ATC CGC     2835
Asp Phe Ser Tyr Asp Leu Thr Leu Ser Tyr Val Pro Asp Leu Ile Arg
    845                 850                 855

AAT GAT CCC AAA TGC ACT ACA GCA CTT GTA ATC AGC GGA GCC TCT TGG     2883
Asn Asp Pro Lys Cys Thr Thr Ala Leu Val Ile Ser Gly Ala Ser Trp
860                 865                 870                 875

GAA ACT TAT GCC AAT AAC TTA GCA CGA CAG GCC TTG CAA GTG CGT GCA     2931
Glu Thr Tyr Ala Asn Asn Leu Ala Arg Gln Ala Leu Gln Val Arg Ala
                880                 885                 890

GGC AGT CAC TAC GCC TTC TCT CCT ATG TTT GAA GTG CTC GGC CAG TTT     2979
Gly Ser His Tyr Ala Phe Ser Pro Met Phe Glu Val Leu Gly Gln Phe
            895                 900                 905

GTC TTT GAA GTT CGT GGA TCC                                        3000
Val Phe Glu Val Arg Gly Ser
        910
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 914 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Lys Ser Gln Phe Ser Trp Leu Val Leu Ser Ser Thr Leu Ala Cys
 1               5                  10                  15
Phe Thr Ser Cys Ser Thr Val Phe Ala Ala Thr Ala Glu Asn Ile Gly
            20                  25                  30
Pro Ser Asp Ser Phe Asp Gly Ser Thr Asn Thr Gly Thr Tyr Thr Pro
        35                  40                  45
Lys Asn Thr Thr Thr Gly Ile Asp Tyr Thr Leu Thr Gly Asp Ile Thr
    50                  55                  60
Leu Gln Asn Leu Gly Asp Ser Ala Ala Leu Thr Lys Gly Cys Phe Ser
65                  70                  75                  80
Asp Thr Thr Glu Ser Leu Ser Phe Ala Gly Lys Gly Tyr Ser Leu Ser
                85                  90                  95
Phe Leu Asn Ile Lys Ser Ser Ala Glu Gly Ala Ala Leu Ser Val Thr
            100                 105                 110
Thr Asp Lys Asn Leu Ser Leu Thr Gly Phe Ser Ser Leu Thr Phe Leu
        115                 120                 125
Ala Ala Pro Ser Ser Val Ile Thr Thr Pro Ser Gly Lys Gly Ala Val
    130                 135                 140
Lys Cys Gly Gly Asp Leu Thr Phe Asp Asn Asn Gly Thr Ile Leu Phe
145                 150                 155                 160
Lys Gln Asp Tyr Cys Glu Glu Asn Gly Gly Ala Ile Ser Thr Lys Asn
                165                 170                 175
Leu Ser Leu Lys Asn Ser Thr Gly Ser Ile Ser Phe Glu Gly Asn Lys
            180                 185                 190
Ser Ser Ala Thr Gly Lys Lys Gly Gly Ala Ile Cys Ala Thr Gly Thr
        195                 200                 205
Val Asp Ile Thr Asn Asn Thr Ala Pro Thr Leu Phe Ser Asn Asn Ile
    210                 215                 220
Ala Glu Ala Ala Gly Gly Ala Ile Asn Ser Thr Gly Asn Cys Thr Ile
225                 230                 235                 240
Thr Gly Asn Thr Ser Leu Val Phe Ser Glu Asn Ser Val Thr Ala Thr
                245                 250                 255
Ala Gly Asn Gly Gly Ala Leu Ser Gly Asp Ala Asp Val Thr Ile Ser
            260                 265                 270
Gly Asn Gln Ser Val Thr Phe Ser Gly Asn Gln Ala Val Ala Asn Gly
        275                 280                 285
Gly Ala Ile Tyr Ala Lys Lys Leu Thr Leu Ala Ser Gly Gly Gly Gly
    290                 295                 300
Gly Ile Ser Phe Ser Asn Asn Ile Val Gln Gly Thr Thr Ala Gly Asn
305                 310                 315                 320
Gly Gly Ala Ile Ser Ile Leu Ala Ala Gly Glu Cys Ser Leu Ser Ala
                325                 330                 335
Glu Ala Gly Asp Ile Thr Phe Asn Gly Asn Ala Ile Val Ala Thr Thr
            340                 345                 350
```

```
Pro Gln Thr Thr Lys Arg Asn Ser Ile Asp Ile Gly Ser Thr Ala Lys
        355                 360                 365
Ile Thr Asn Leu Arg Ala Ile Ser Gly His Ser Ile Phe Phe Tyr Asp
    370                 375                 380
Pro Ile Thr Ala Asn Thr Ala Ala Asp Ser Thr Asp Thr Leu Asn Leu
385                 390                 395                 400
Asn Lys Ala Asp Ala Gly Asn Ser Thr Asp Tyr Ser Gly Ser Ile Val
                405                 410                 415
Phe Ser Gly Glu Lys Leu Ser Glu Asp Glu Ala Lys Val Ala Asp Asn
            420                 425                 430
Leu Thr Ser Thr Leu Lys Gln Pro Val Thr Leu Thr Ala Gly Asn Leu
        435                 440                 445
Val Leu Lys Arg Gly Val Thr Leu Asp Thr Lys Gly Phe Thr Gln Thr
    450                 455                 460
Ala Gly Ser Ser Val Ile Met Asp Ala Gly Thr Thr Leu Lys Ala Ser
465                 470                 475                 480
Thr Glu Glu Val Thr Leu Thr Gly Leu Ser Ile Pro Val Asp Ser Leu
                485                 490                 495
Gly Glu Gly Lys Lys Val Val Ile Ala Ala Ser Ala Ala Ser Lys Asn
            500                 505                 510
Val Ala Leu Ser Gly Pro Ile Leu Leu Leu Asp Asn Gln Gly Asn Ala
        515                 520                 525
Tyr Glu Asn His Asp Leu Gly Lys Thr Gln Asp Phe Ser Phe Val Gln
    530                 535                 540
Leu Ser Ala Leu Gly Thr Ala Thr Thr Thr Asp Val Pro Ala Val Pro
545                 550                 555                 560
Thr Val Ala Thr Pro Thr His Tyr Gly Tyr Gln Gly Thr Trp Gly Met
                565                 570                 575
Thr Trp Val Asp Asp Thr Ala Ser Thr Pro Lys Thr Lys Thr Ala Thr
            580                 585                 590
Leu Ala Trp Thr Asn Thr Gly Tyr Leu Pro Asn Pro Glu Arg Gln Gly
        595                 600                 605
Pro Leu Val Pro Asn Ser Leu Trp Gly Ser Phe Ser Asp Ile Gln Ala
    610                 615                 620
Ile Gln Gly Val Ile Glu Arg Ser Ala Leu Thr Leu Cys Ser Asp Arg
625                 630                 635                 640
Gly Phe Trp Ala Ala Gly Val Ala Asn Phe Leu Asp Lys Asp Lys Lys
                645                 650                 655
Gly Glu Lys Arg Lys Tyr Arg His Lys Ser Gly Gly Tyr Ala Ile Gly
            660                 665                 670
Gly Ala Ala Gln Thr Cys Ser Glu Asn Leu Ile Ser Phe Ala Phe Cys
        675                 680                 685
Gln Leu Phe Gly Ser Asp Lys Asp Phe Leu Val Ala Lys Asn His Thr
    690                 695                 700
Asp Thr Tyr Ala Gly Ala Phe Tyr Ile Gln His Ile Thr Glu Cys Ser
705                 710                 715                 720
Gly Phe Ile Gly Cys Leu Leu Asp Lys Leu Pro Gly Ser Trp Ser His
                725                 730                 735
Lys Pro Leu Val Leu Glu Gly Gln Leu Ala Tyr Ser His Val Ser Asn
            740                 745                 750
Asp Leu Lys Thr Lys Tyr Thr Ala Tyr Pro Glu Val Lys Gly Ser Trp
        755                 760                 765
Gly Asn Asn Ala Phe Asn Met Met Leu Gly Ala Ser Ser His Ser Tyr
    770                 775                 780
Pro Glu Tyr Leu His Cys Phe Asp Thr Tyr Ala Pro Tyr Ile Lys Leu
785                 790                 795                 800
Asn Leu Thr Tyr Ile Arg Gln Asp Ser Phe Ser Glu Lys Gly Thr Glu
                805                 810                 815
```

```
Gly Arg Ser Phe Asp Asp Ser Asn Leu Phe Asn Leu Ser Leu Pro Ile
            820                 825                 830
Gly Val Lys Phe Glu Lys Phe Ser Asp Cys Asn Asp Phe Ser Tyr Asp
        835                 840                 845
Leu Thr Leu Ser Tyr Val Pro Asp Leu Ile Arg Asn Asp Pro Lys Cys
    850                 855                 860
Thr Thr Ala Leu Val Ile Ser Gly Ala Ser Trp Glu Thr Tyr Ala Asn
865                 870                 875                 880
Asn Leu Ala Arg Gln Ala Leu Gln Val Arg Ala Gly Ser His Tyr Ala
                885                 890                 895
Phe Ser Pro Met Phe Glu Val Leu Gly Gln Phe Val Phe Glu Val Arg
            900                 905                 910
Gly Ser
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1200 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...1200
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
GAT CCT AAA AAT AAA GAG TAC ACA GGG ACC ATA CTC TTT TCT GGA GAA      48
Asp Pro Lys Asn Lys Glu Tyr Thr Gly Thr Ile Leu Phe Ser Gly Glu
 1               5                  10                  15

AAG AGT CTA GCA AAC GAT CCT AGG GAT TTT AAA TCT ACA ATC CCT CAG      96
Lys Ser Leu Ala Asn Asp Pro Arg Asp Phe Lys Ser Thr Ile Pro Gln
            20                  25                  30

AAC GTC AAC CTG TCT GCA GGA TAC TTA GTT ATT AAA GAG GGG GCC GAA     144
Asn Val Asn Leu Ser Ala Gly Tyr Leu Val Ile Lys Glu Gly Ala Glu
        35                  40                  45

GTC ACA GTT TCA AAA TTC ACG CAG TCT CCA GGA TCG CAT TTA GTT TTA     192
Val Thr Val Ser Lys Phe Thr Gln Ser Pro Gly Ser His Leu Val Leu
    50                  55                  60

GAT TTA GGA ACC AAA CTG ATA GCC TCT AAG GAA GAC ATT GCC ATC ACA     240
Asp Leu Gly Thr Lys Leu Ile Ala Ser Lys Glu Asp Ile Ala Ile Thr
65                  70                  75                  80

GGC CTC GCG ATA GAT ATA GAT AGC TTA AGC TCA TCC TCA ACA GCA GCT     288
Gly Leu Ala Ile Asp Ile Asp Ser Leu Ser Ser Ser Ser Thr Ala Ala
                85                  90                  95
```

```
GTT ATT AAA GCA AAC ACC GCA AAT AAA CAG ATA TCC GTG ACG GAC TCT     336
Val Ile Lys Ala Asn Thr Ala Asn Lys Gln Ile Ser Val Thr Asp Ser
            100                 105                 110

ATA GAA CTT ATC TCG CCT ACT GGC AAT GCC TAT GAA GAT CTC AGA ATG     384
Ile Glu Leu Ile Ser Pro Thr Gly Asn Ala Tyr Glu Asp Leu Arg Met
        115                 120                 125

AGA AAT TCA CAG ACG TTC CCT CTG CTC TCT TTA GAG CCT GGA GCC GGG     432
Arg Asn Ser Gln Thr Phe Pro Leu Leu Ser Leu Glu Pro Gly Ala Gly
    130                 135                 140

GGT AGT GTG ACT GTN ACT GCT GGA GAT TTC CTA CCG GTA AGT CCC CAT     480
Gly Ser Val Thr Val Thr Ala Gly Asp Phe Leu Pro Val Ser Pro His
145                 150                 155                 160

TAT GGT TTT CAA GGC AAT TGG AAA TTA GCT TGG ACA GGA ACT GGA AAC     528
Tyr Gly Phe Gln Gly Asn Trp Lys Leu Ala Trp Thr Gly Thr Gly Asn
                165                 170                 175

AAA GTT GGA GAA TTC TTC TGG GAT AAA ATA AAT TAT AAG CCT AGA CCT     576
Lys Val Gly Glu Phe Phe Trp Asp Lys Ile Asn Tyr Lys Pro Arg Pro
            180                 185                 190

GAA AAA GAA GGA AAT TTA GTT CCT AAT ATC TTG TGG GGG AAT GCT GTA     624
Glu Lys Glu Gly Asn Leu Val Pro Asn Ile Leu Trp Gly Asn Ala Val
        195                 200                 205

AAT GTC AGA TCC TTA ATG CAG GTT CAA GAG ACC CAT GCA TCG AGC TTA     672
Asn Val Arg Ser Leu Met Gln Val Gln Glu Thr His Ala Ser Ser Leu
    210                 215                 220

CAG ACA GAT CGA GGG CTG TGG ATC GAT GGA ATT GGG AAT TTC TTC CAT     720
Gln Thr Asp Arg Gly Leu Trp Ile Asp Gly Ile Gly Asn Phe Phe His
225                 230                 235                 240

GTA TCT GCC TCC GAA GAC AAT ATA AGG TAC CGT CAT AAC AGC GGT GGA     768
Val Ser Ala Ser Glu Asp Asn Ile Arg Tyr Arg His Asn Ser Gly Gly
                245                 250                 255

TAT GTT CTA TCT GTA AAT AAT GAG ATC ACA CCT AAG CAC TAT ACT TCG     816
Tyr Val Leu Ser Val Asn Asn Glu Ile Thr Pro Lys His Tyr Thr Ser
            260                 265                 270

ATG GCA TTT TCC CAA CTC TTT AGT AGA GAC AAN GAC TAT GCG GTT TCC     864
Met Ala Phe Ser Gln Leu Phe Ser Arg Asp Lys Asp Tyr Ala Val Ser
        275                 280                 285

AAC AAC GAA TAC AGA ATG TAT TTA GGA TCG TAT CTC TAT CAA TAT ACA     912
Asn Asn Glu Tyr Arg Met Tyr Leu Gly Ser Tyr Leu Tyr Gln Tyr Thr
    290                 295                 300

ACC TCC CTA GGG AAT ATT TTC CGT TAT GCT TCG CGT AAC CCT AAT GTA     960
Thr Ser Leu Gly Asn Ile Phe Arg Tyr Ala Ser Arg Asn Pro Asn Val
305                 310                 315                 320
```

```
AAC GTC GGG ATT CTC TCA AGA AGG TTT CTT CAA AAT CCT CTT ATG ATT    1008
Asn Val Gly Ile Leu Ser Arg Arg Phe Leu Gln Asn Pro Leu Met Ile
                325                 330                 335

TTT CAT TTT TTG TGT GCT TAT GGT CAT GCC ACC AAT GAT ATG AAA ACA    1056
Phe His Phe Leu Cys Ala Tyr Gly His Ala Thr Asn Asp Met Lys Thr
            340                 345                 350

GAC TAC GCN AAT TTC CCT ATG GTG AAA AAC AGC TGG AGA AAC AAT TGT    1104
Asp Tyr Ala Asn Phe Pro Met Val Lys Asn Ser Trp Arg Asn Asn Cys
        355                 360                 365

TGG GCT NTA AAN TGC GGA GGG AGC ATG CCT CTA TTG GTA TTT GAN AAC    1152
Trp Ala Ile Lys Cys Gly Gly Ser Met Pro Leu Leu Val Phe Glu Asn
    370                 375                 380

GGA ANA CTT TTC CAA GGT GCC ATC CCN TTT ATG AAA CTA CAA TTA GTT    1200
Gly Lys Leu Phe Gln Gly Ala Ile Pro Phe Met Lys Leu Gln Leu Val
385                 390                 395                 400
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 400 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Asp Pro Lys Asn Lys Glu Tyr Thr Gly Thr Ile Leu Phe Ser Gly Glu
 1               5                  10                  15
Lys Ser Leu Ala Asn Asp Pro Arg Asp Phe Lys Ser Thr Ile Pro Gln
            20                  25                  30
Asn Val Asn Leu Ser Ala Gly Tyr Leu Val Ile Lys Glu Gly Ala Glu
        35                  40                  45
Val Thr Val Ser Lys Phe Thr Gln Ser Pro Gly Ser His Leu Val Leu
    50                  55                  60
Asp Leu Gly Thr Lys Leu Ile Ala Ser Lys Glu Asp Ile Ala Ile Thr
65                  70                  75                  80
Gly Leu Ala Ile Asp Ile Asp Ser Leu Ser Ser Ser Ser Thr Ala Ala
                85                  90                  95
Val Ile Lys Ala Asn Thr Ala Asn Lys Gln Ile Ser Val Thr Asp Ser
            100                 105                 110
Ile Glu Leu Ile Ser Pro Thr Gly Asn Ala Tyr Glu Asp Leu Arg Met
        115                 120                 125
Arg Asn Ser Gln Thr Phe Pro Leu Leu Ser Leu Glu Pro Gly Ala Gly
    130                 135                 140
Gly Ser Val Thr Val Thr Ala Gly Asp Phe Leu Pro Val Ser Pro His
145                 150                 155                 160
Tyr Gly Phe Gln Gly Asn Trp Lys Leu Ala Trp Thr Gly Thr Gly Asn
                165                 170                 175
```

```
Lys Val Gly Glu Phe Phe Trp Asp Lys Ile Asn Tyr Lys Pro Arg Pro
            180                 185                 190
Glu Lys Glu Gly Asn Leu Val Pro Asn Ile Leu Trp Gly Asn Ala Val
        195                 200                 205
Asn Val Arg Ser Leu Met Gln Val Gln Glu Thr His Ala Ser Ser Leu
    210                 215                 220
Gln Thr Asp Arg Gly Leu Trp Ile Asp Gly Ile Gly Asn Phe Phe His
225                 230                 235                 240
Val Ser Ala Ser Glu Asp Asn Ile Arg Tyr Arg His Asn Ser Gly Gly
                245                 250                 255
Tyr Val Leu Ser Val Asn Asn Glu Ile Thr Pro Lys His Tyr Thr Ser
            260                 265                 270
Met Ala Phe Ser Gln Leu Phe Ser Arg Asp Lys Asp Tyr Ala Val Ser
        275                 280                 285
Asn Asn Glu Tyr Arg Met Tyr Leu Gly Ser Tyr Leu Tyr Gln Tyr Thr
    290                 295                 300
Thr Ser Leu Gly Asn Ile Phe Arg Tyr Ala Ser Arg Asn Pro Asn Val
305                 310                 315                 320
Asn Val Gly Ile Leu Ser Arg Arg Phe Leu Gln Asn Pro Leu Met Ile
                325                 330                 335
Phe His Phe Leu Cys Ala Tyr Gly His Ala Thr Asn Asp Met Lys Thr
            340                 345                 350
Asp Tyr Ala Asn Phe Pro Met Val Lys Asn Ser Trp Arg Asn Asn Cys
        355                 360                 365
Trp Ala Ile Lys Cys Gly Gly Ser Met Pro Leu Leu Val Phe Glu Asn
    370                 375                 380
Gly Lys Leu Phe Gln Gly Ala Ile Pro Phe Met Lys Leu Gln Leu Val
385                 390                 395                 400
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1830 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...1830
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GAT CTC ACA TTA GGG AGT CGT GAC AGT TAT AAT GGT GAT ACA AGC ACC      48
Asp Leu Thr Leu Gly Ser Arg Asp Ser Tyr Asn Gly Asp Thr Ser Thr
 1               5                  10                  15

ACA GAA TTT ACT CCT AAA GCG GCA ACT TCT GAT GCT AGT GGC ACG ACC      96
Thr Glu Phe Thr Pro Lys Ala Ala Thr Ser Asp Ala Ser Gly Thr Thr
            20                  25                  30

TAT ATT CTC GAT GGG GAT GTC TCG ATA AGC CAA GCA GGG AAA CAA ACG     144
Tyr Ile Leu Asp Gly Asp Val Ser Ile Ser Gln Ala Gly Lys Gln Thr
        35                  40                  45
```

```
AGC TTA ACC ACA AGT TGT TTT TCT AAC ACT GCA GGA AAT CTT ACC TTC      192
Ser Leu Thr Thr Ser Cys Phe Ser Asn Thr Ala Gly Asn Leu Thr Phe
    50                  55                  60

TTA GGG AAC GGA TTT TCT CTT CAT TTT GAC AAT ATT ATT TCG TCT ACT      240
Leu Gly Asn Gly Phe Ser Leu His Phe Asp Asn Ile Ile Ser Ser Thr
65                  70                  75                  80

GTT GCA GGT GTT GTT GTT AGC AAT ACA GCA GCT TCT GGG ATT ACG AAA      288
Val Ala Gly Val Val Val Ser Asn Thr Ala Ala Ser Gly Ile Thr Lys
                85                  90                  95

TTC TCA GGA TTT TCA ACT CTT CGG ATG CTT GCA GCT CCT AGG ACC ACA      336
Phe Ser Gly Phe Ser Thr Leu Arg Met Leu Ala Ala Pro Arg Thr Thr
            100                 105                 110


GGT AAA GGA GCC ATT AAA ATT ACC GAT GGT CTG GTG TTT GAG AGT ATA      384
Gly Lys Gly Ala Ile Lys Ile Thr Asp Gly Leu Val Phe Glu Ser Ile
        115                 120                 125

GGG AAT CTT GAT CCG ATT ACT GTA ACA GGA TCG ACA TCT GTT GCT GAT      432
Gly Asn Leu Asp Pro Ile Thr Val Thr Gly Ser Thr Ser Val Ala Asp
    130                 135                 140

GCT CTC AAT ATT AAT AGC CCT GAT ACT GGA GAT AAC AAA GAG TAT ACG      480
Ala Leu Asn Ile Asn Ser Pro Asp Thr Gly Asp Asn Lys Glu Tyr Thr
145                 150                 155                 160

GGA ACC ATA GTC TTT TCT GGA GAG AAG CTC ACG GAG GCA GAA GCT AAA      528
Gly Thr Ile Val Phe Ser Gly Glu Lys Leu Thr Glu Ala Glu Ala Lys
                165                 170                 175

GAT GAG AAG AAC CGC ACT TCT AAA TTA CTT CAA AAT GTT GCT TTT AAA      576
Asp Glu Lys Asn Arg Thr Ser Lys Leu Leu Gln Asn Val Ala Phe Lys
            180                 185                 190

AAT GGG ACT GTA GTT TTA AAA GGT GAT GTC GTT TTA AGT GCG AAC GGT      624
Asn Gly Thr Val Val Leu Lys Gly Asp Val Val Leu Ser Ala Asn Gly
        195                 200                 205

TTC TCT CAG GAT GCA AAC TCT AAG TTG ATT ATG GAT TTA GGG ACG TCG      672
Phe Ser Gln Asp Ala Asn Ser Lys Leu Ile Met Asp Leu Gly Thr Ser
    210                 215                 220

TTG GTT GCA AAC ACC GAA AGT ATC GAG TTA ACG AAT TTG GAA ATT AAT      720
Leu Val Ala Asn Thr Glu Ser Ile Glu Leu Thr Asn Leu Glu Ile Asn
225                 230                 235                 240

ATA GAC TCT CTC AGG AAC GGG AAA AAG ATA AAA CTC AGT GCT GCC ACA      768
Ile Asp Ser Leu Arg Asn Gly Lys Lys Ile Lys Leu Ser Ala Ala Thr
                245                 250                 255

GCT CAG AAA GAT ATT CGT ATA GAT CGT CCT GTT GTA CTG GCA ATT AGC      816
Ala Gln Lys Asp Ile Arg Ile Asp Arg Pro Val Val Leu Ala Ile Ser
            260                 265                 270
```

```
GAT GAG AGT TTT TAT CAA AAT GGC TTT TTG AAT GAG GAC CAT TCC TAT      864
Asp Glu Ser Phe Tyr Gln Asn Gly Phe Leu Asn Glu Asp His Ser Tyr
        275                 280                 285

GAT GGG ATT CTT GAG TTA GAT GCT GGG AAA GAC ATC GTG ATT TCT GCA      912
Asp Gly Ile Leu Glu Leu Asp Ala Gly Lys Asp Ile Val Ile Ser Ala
    290                 295                 300

GAT TCT CGC AGT ATA GAT GCT GTA CAA TCT CCG TAT GGC TAT CAG GGA      960
Asp Ser Arg Ser Ile Asp Ala Val Gln Ser Pro Tyr Gly Tyr Gln Gly
305                 310                 315                 320

AAG TGG ACG ATC AAT TGG TCT ACT GAT GAT AAG AAA GCT ACG GTT TCT     1008
Lys Trp Thr Ile Asn Trp Ser Thr Asp Asp Lys Lys Ala Thr Val Ser
                325                 330                 335

TGG GCG AAG CAG AGT TTT AAT CCC ACT GCT GAG CAG GAG GCT CCG TTA     1056
Trp Ala Lys Gln Ser Phe Asn Pro Thr Ala Glu Gln Glu Ala Pro Leu
            340                 345                 350

GTT CCT AAT CTT CTT TGG GGT TCT TTT ATA GAT GTT CGT TCC TTC CAG     1104
Val Pro Asn Leu Leu Trp Gly Ser Phe Ile Asp Val Arg Ser Phe Gln
        355                 360                 365

AAT TTT ATA GAG CTA GGT ACT GAA GGT GCT CCT TAC GAA AAG AGA TTT     1152
Asn Phe Ile Glu Leu Gly Thr Glu Gly Ala Pro Tyr Glu Lys Arg Phe
    370                 375                 380

TGG GTT GCA GGC ATT TCC AAT GTT TTG CAT AGG AGC GGT CGT GAA AAT     1200
Trp Val Ala Gly Ile Ser Asn Val Leu His Arg Ser Gly Arg Glu Asn
385                 390                 395                 400

CAA AGG AAA TTC CGT CAT GTG AGT GGA GGT GCT GTA GTA GGT GCT AGC     1248
Gln Arg Lys Phe Arg His Val Ser Gly Gly Ala Val Val Gly Ala Ser
                405                 410                 415

ACG AGG ATG CCG GGT GGT GAT ACC TTG TCT CTG GGT TTT GCT CAG CTC     1296
Thr Arg Met Pro Gly Gly Asp Thr Leu Ser Leu Gly Phe Ala Gln Leu
            420                 425                 430

TTT GCG CGT GAC AAA GAC TAC TTT ATG AAT ACC AAT TTC GCA AAG ACC     1344
Phe Ala Arg Asp Lys Asp Tyr Phe Met Asn Thr Asn Phe Ala Lys Thr
        435                 440                 445

TAC GCA GGA TCT TTA CGT TTG CAG CAC GAT GCT TCC CTA TAC TCT GTG     1392
Tyr Ala Gly Ser Leu Arg Leu Gln His Asp Ala Ser Leu Tyr Ser Val
    450                 455                 460

GTG AGT ATC CTT TTA GGA GAG GGA GGA CTC CGC GAG ATC CTG TTG CCT     1440
Val Ser Ile Leu Leu Gly Glu Gly Gly Leu Arg Glu Ile Leu Leu Pro
465                 470                 475                 480

TAT GTT TCC AAT ACT CTG CCG TGC TCT TTC TAT GGG CAG CTT AGC TAC     1488
Tyr Val Ser Asn Thr Leu Pro Cys Ser Phe Tyr Gly Gln Leu Ser Tyr
                485                 490                 495
```

```
GGC CAT ACG GAT CAT CGC ATG AAG ACC GAG TCT CTA CCC CCC CCC CCC    1536
Gly His Thr Asp His Arg Met Lys Thr Glu Ser Leu Pro Pro Pro Pro
            500                 505                 510

CCG ACG CTC TCG ACG GAT CAT ACT TCT TGG GGA GGA TAT GTC TGG GCT    1584
Pro Thr Leu Ser Thr Asp His Thr Ser Trp Gly Gly Tyr Val Trp Ala
        515                 520                 525

GGA GAG CTG GGA ACT CGA GTT GCT GTT GAA AAT ACC AGC GGC AGA GGA    1632
Gly Glu Leu Gly Thr Arg Val Ala Val Glu Asn Thr Ser Gly Arg Gly
    530                 535                 540

TTT TTC CGA GAG TAC ACT CCA TTT GTA AAA GTC CAA GCT GTT TAC TCG    1680
Phe Phe Arg Glu Tyr Thr Pro Phe Val Lys Val Gln Ala Val Tyr Ser
545                 550                 555                 560

CGC CAA GAT AGC TTT GTT GAA CTA GGA GCT ATC AGT CGT GAT TTT AGT    1728
Arg Gln Asp Ser Phe Val Glu Leu Gly Ala Ile Ser Arg Asp Phe Ser
                565                 570                 575

GAT TCG CAT CTT TAT AAC CTT GCG ATT CCT CTT GGA ATC AAG TTA GAG    1776
Asp Ser His Leu Tyr Asn Leu Ala Ile Pro Leu Gly Ile Lys Leu Glu
            580                 585                 590

AAA CGG TTT GCA GAG CAA TAT TAT CAT GTT GTT GCG ATG TAT TCT CCA    1824
Lys Arg Phe Ala Glu Gln Tyr Tyr His Val Val Ala Met Tyr Ser Pro
        595                 600                 605

GAT GTT                                                            1830
Asp Val
    610
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 610 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Asp Leu Thr Leu Gly Ser Arg Asp Ser Tyr Asn Gly Asp Thr Ser Thr
 1               5                  10                  15
Thr Glu Phe Thr Pro Lys Ala Ala Thr Ser Asp Ala Ser Gly Thr Thr
            20                  25                  30
Tyr Ile Leu Asp Gly Asp Val Ser Ile Ser Gln Ala Gly Lys Gln Thr
        35                  40                  45
Ser Leu Thr Thr Ser Cys Phe Ser Asn Thr Ala Gly Asn Leu Thr Phe
    50                  55                  60
Leu Gly Asn Gly Phe Ser Leu His Phe Asp Asn Ile Ile Ser Ser Thr
65                  70                  75                  80
Val Ala Gly Val Val Val Ser Asn Thr Ala Ala Ser Gly Ile Thr Lys
                85                  90                  95
```

```
Phe Ser Gly Phe Ser Thr Leu Arg Met Leu Ala Ala Pro Arg Thr Thr
            100                 105                 110
Gly Lys Gly Ala Ile Lys Ile Thr Asp Gly Leu Val Phe Glu Ser Ile
        115                 120                 125
Gly Asn Leu Asp Pro Ile Thr Val Thr Gly Ser Thr Ser Val Ala Asp
    130                 135                 140
Ala Leu Asn Ile Asn Ser Pro Asp Thr Gly Asp Asn Lys Glu Tyr Thr
145                 150                 155                 160
Gly Thr Ile Val Phe Ser Gly Glu Lys Leu Thr Glu Ala Glu Ala Lys
                165                 170                 175
Asp Glu Lys Asn Arg Thr Ser Lys Leu Leu Gln Asn Val Ala Phe Lys
            180                 185                 190
Asn Gly Thr Val Val Leu Lys Gly Asp Val Val Leu Ser Ala Asn Gly
        195                 200                 205
Phe Ser Gln Asp Ala Asn Ser Lys Leu Ile Met Asp Leu Gly Thr Ser
    210                 215                 220
Leu Val Ala Asn Thr Glu Ser Ile Glu Leu Thr Asn Leu Glu Ile Asn
225                 230                 235                 240
Ile Asp Ser Leu Arg Asn Gly Lys Lys Ile Lys Leu Ser Ala Ala Thr
                245                 250                 255
Ala Gln Lys Asp Ile Arg Ile Asp Arg Pro Val Val Leu Ala Ile Ser
            260                 265                 270
Asp Glu Ser Phe Tyr Gln Asn Gly Phe Leu Asn Glu Asp His Ser Tyr
        275                 280                 285
Asp Gly Ile Leu Glu Leu Asp Ala Gly Lys Asp Ile Val Ile Ser Ala
    290                 295                 300
Asp Ser Arg Ser Ile Asp Ala Val Gln Ser Pro Tyr Gly Tyr Gln Gly
305                 310                 315                 320
Lys Trp Thr Ile Asn Trp Ser Thr Asp Asp Lys Lys Ala Thr Val Ser
                325                 330                 335
Trp Ala Lys Gln Ser Phe Asn Pro Thr Ala Glu Gln Glu Ala Pro Leu
            340                 345                 350
Val Pro Asn Leu Leu Trp Gly Ser Phe Ile Asp Val Arg Ser Phe Gln
        355                 360                 365
Asn Phe Ile Glu Leu Gly Thr Glu Gly Ala Pro Tyr Glu Lys Arg Phe
    370                 375                 380
Trp Val Ala Gly Ile Ser Asn Val Leu His Arg Ser Gly Arg Glu Asn
385                 390                 395                 400
Gln Arg Lys Phe Arg His Val Ser Gly Gly Ala Val Val Gly Ala Ser
                405                 410                 415
Thr Arg Met Pro Gly Gly Asp Thr Leu Ser Leu Gly Phe Ala Gln Leu
            420                 425                 430
Phe Ala Arg Asp Lys Asp Tyr Phe Met Asn Thr Asn Phe Ala Lys Thr
        435                 440                 445
Tyr Ala Gly Ser Leu Arg Leu Gln His Asp Ala Ser Leu Tyr Ser Val
    450                 455                 460
Val Ser Ile Leu Leu Gly Glu Gly Gly Leu Arg Glu Ile Leu Leu Pro
465                 470                 475                 480
Tyr Val Ser Asn Thr Leu Pro Cys Ser Phe Tyr Gly Gln Leu Ser Tyr
                485                 490                 495
Gly His Thr Asp His Arg Met Lys Thr Glu Ser Leu Pro Pro Pro Pro
            500                 505                 510
Pro Thr Leu Ser Thr Asp His Thr Ser Trp Gly Gly Tyr Val Trp Ala
        515                 520                 525
Gly Glu Leu Gly Thr Arg Val Ala Val Glu Asn Thr Ser Gly Arg Gly
    530                 535                 540
Phe Phe Arg Glu Tyr Thr Pro Phe Val Lys Val Gln Ala Val Tyr Ser
545                 550                 555                 560
```

```
Arg Gln Asp Ser Phe Val Glu Leu Gly Ala Ile Ser Arg Asp Phe Ser
                565                 570                 575
Asp Ser His Leu Tyr Asn Leu Ala Ile Pro Leu Gly Ile Lys Leu Glu
            580                 585                 590
Lys Arg Phe Ala Glu Gln Tyr Tyr His Val Val Ala Met Tyr Ser Pro
        595                 600                 605
Asp Val
    610
```

Claims:

1. Species specific diagnostic test for infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said test comprising detecting in a patient sample the presence of antibodies against proteins from the outer membrane of *Clamydia pneumoniae*, said proteins being of a molecular weight of 98/95 kDa, or detecting the presence of nucleic acid fragments encoding said outer membrane proteins.

2. Diagnostic test according to claim 1, wherein the outer membrane protein has the sequence shown in SEQ ID NO: 2, or a variant or subsequence thereof.

3. Diagnostic test according to claim 1, wherein the outer membrane protein has the sequence shown in SEQ ID NO: 4, or a variant or subsequence thereof.

4. Diagnostic test according to claim 1, wherein the outer membrane protein has the sequence shown in SEQ ID NO: 6, or a variant or subsequence thereof.

5. Diagnostic test according to claim 1, wherein the outer membrane protein has the sequence shown in SEQ ID NO: 8, or a variant or subsequence thereof.

6. Diagnostic test according to claim 1, wherein the nucleic acid fragment has the sequence shown in SEQ ID NO: 1 or a variant or subsequence thereof.

7. Diagnostic test according to claim 1, wherein the nucleic acid fragment has the sequence shown in SEQ ID NO: 3 or a variant or subsequence thereof.

8. Diagnostic test according to claim 1, wherein the nucleic acid fragment has the sequence shown in SEQ ID NO: 5 or a variant or subsequence thereof.

9. Diagnostic test according to claim 1, wherein the nucleic acid fragment has the sequence shown in SEQ ID NO: 7 or a variant or subsequence thereof.

10. Diagnostic test according to any of claims 6-9, wherein detection of nucleic acid fragments is obtained by using nucleic acid amplification.

11. Diagnostic test according to claim 10, wherein detection of nucleic acid fragments is obtained by using polymerase chain reaction.

12. A nucleic acid fragment derived from *Chlamydia pneumoniae* comprising the nucleotide sequence SEQ ID NO: 1, or a variant or subsequence of said nucleotide sequence which has a sequence homology of at least 50% with SEQ ID NO: 1.

13. A nucleic acid fragment derived from *Chlamydia pneumoniae* comprising the nucleotide sequence SEQ ID NO: 3, or a variant or subsequence of said nucleotide sequence which has a sequence homology of at least 50% with SEQ ID NO: 3.

14. A nucleic acid fragment derived from *Chlamydia pneumoniae* comprising the nucleotide sequence SEQ ID NO: 5, or a variant or subsequence of said nucleotide sequence which has a sequence homology of at least 50% with SEQ ID NO:5.

15. A nucleic acid fragment derived from *Chlamydia pneumoniae* comprising the nucleotide sequence SEQ ID NO: 7, or a variant or subsequence of said nucleotide sequence which has a sequence homology of at least 50% with SEQ ID NO: 7.

16. An at least partially purified nucleic acid fragment accoring to any of claims 12-15.

17. A protein derived from *Chlamydia pneumoniae* having the amino acid sequence shown in SEQ ID NO: 2 or variant or

subsequence thereof having a sequence homology of at least 50% and a substantially identical biological activity.

18. A protein derived from *Chlamydia pneumoniae* having the amino acid sequence shown in SEQ ID NO: 4 or variant or subsequence thereof having a sequence homology of at least 50% and a substantially identical biological activity.

19. A protein derived from *Chlamydia pneumoniae* having the amino acid sequence shown in SEQ ID NO: 6 or variant or subsequence thereof having a sequence homology of at least 50% and a substantially identical biological activity.

20. A protein derived from *Chlamydia pneumoniae* having the amino acid sequence shown in SEQ ID NO: 8 or variant or subsequence thereof having a sequence homology of at least 50% and a substantially identical biological activity.

21. Polyclonal monospecific antibody against the protein with the sequence shown in SEQ ID NO: 2 or a variant or subsequence thereof.

22. Polyclonal monospecific antibody against the protein with the sequence shown in SEQ ID NO: 4 or a variant or subsequence thereof.

23. Polyclonal monospecific antibody against the protein with the sequence shown in SEQ ID NO: 6 or a variant or subsequence thereof.

24. Polyclonal monospecific antibody against the protein with the sequence shown in SEQ ID NO: 8 or a variant or subsequence thereof.

25. Monoclonal antibody against the protein with the sequence shown in SEQ ID NO: 2 or a variant or subsequence thereof.

26. Monoclonal antibody against the protein with the sequence shown in SEQ ID NO: 4 or a variant or subsequence thereof.

27. Monoclonal antibody against the protein with the sequence shown in SEQ ID NO: 6 or a variant or subsequence thereof.

28. Monoclonal antibody against the protein with the sequence shown in SEQ ID NO: 8 or a variant or subsequence thereof.

29. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a protein with the amino acid sequence SEQ ID NO: 2, or a variant or subsequence thereof.

30. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a protein with the amino acid sequence SEQ ID NO: 4, or a variant or subsequence thereof.

31. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a protein with the amino acid sequence SEQ ID NO: 6, or a variant or subsequence thereof.

32. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a protein with the amino acid sequence SEQ ID NO: 8, or a variant or subsequence thereof.

33. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising antibodies against a protein with the amino acid sequence SEQ ID NO: 2, or a variant or subsequence thereof.

34. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising antibodies against a protein with the amino acid sequence SEQ ID NO: 4, or a variant or subsequence thereof.

35. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising antibodies against a protein with the amino acid sequence SEQ ID NO: 6, or a variant or subsequence thereof.

36. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising antibodies against a protein with the amino acid sequence SEQ ID NO: 8, or a variant or subsequence thereof.

37. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a nucleic acid fragment with the sequence SEQ ID NO: 1, or a variant or subsequence thereof.

38. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a nucleic acid fragment with the sequence SEQ ID NO: 3, or a variant or subsequence thereof.

39. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a nucleic acid fragment with the sequence SEQ ID NO: 5, or a variant or subsequence thereof.

40. A diagnostic kit for the diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*, said kit comprising a nucleic acid fragment with the sequence SEQ ID NO: 7, or a variant or subsequence thereof.

41. A composition for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*, said composition

comprising a protein with the amino acid sequence shown in SEQ ID NO: 2, or a variant or subsequence thereof.

42. A composition for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*, said composition comprising a protein with the amino acid sequence shown in SEQ ID NO: 4, or a variant or subsequence thereof.

43. A composition for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*, said composition comprising a protein with the amino acid sequence shown in SEQ ID NO: 6, or a variant or subsequence thereof.

44. A composition for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*, said composition comprising a protein with the amino acid sequence shown in SEQ ID NO: 8, or a variant or subsequence thereof.

45. Use of a protein with the sequence shown in SEQ ID NO: 2 or a variant or subsequence thereof in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

46. Use of a protein with the sequence shown in SEQ ID NO: 4 or a variant or subsequence thereof in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

47. Use of a protein with the sequence shown in SEQ ID NO: 6 or a variant or subsequence thereof in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

48. Use of a protein with the sequence shown in SEQ ID NO: 8 or a variant or subsequence thereof in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

49. Use of the protein with the sequence shown in SEQ ID NO: 2 or a variant or subsequence thereof in an undenatured form, in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

50. Use of the protein with the sequence shown in SEQ ID NO: 4 or a variant or subsequence thereof in an undenatured form, in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

51. Use of the protein with the sequence shown in SEQ ID NO: 6 or a variant or subsequence thereof in an undenatured form, in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

52. Use of the protein with the sequence shown in SEQ ID NO: 8 or a variant or subsequence thereof in an undenatured form, in diagnosis of infection of a mammal, such as a human, with *Chlamydia pneumoniae*.

53. Use of a protein with the sequence shown in SEQ ID NO: 2, or a variant or subsequence thereof, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

54. Use of a protein with the sequence shown in SEQ ID NO: 4, or a variant or subsequence thereof, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

55. Use of a protein with the sequence shown in SEQ ID NO: 6, or a variant or subsequence thereof, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

56. Use of a protein with the sequence shown in SEQ ID NO: 8, or a variant or subsequence thereof, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

57. Use of the protein with the sequence shown in SEQ ID NO: 2 or a variant or subsequence thereof in an undenatured

form, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

58. Use of the protein with the sequence shown in SEQ ID NO: 4 or a variant or subsequence thereof in an undenatured form, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

59. Use of the protein with the sequence shown in SEQ ID NO: 6 or a variant or subsequence thereof in an undenatured form, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

60. Use of the protein with the sequence shown in SEQ ID NO: 8 or a variant or subsequence thereof in an undenatured form, for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

61. Use of a nucleic acid fragment with the nucleotide sequence shown in SEQ ID NO: 1 or a variant or subsequence of said nucleotide sequence showing a sequence homology of at least 50% to SEQ ID NO: 1 for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

62. Use of a nucleic acid fragment with the nucleotide sequence shown in SEQ ID NO: 3 or a variant or subsequence of said nucleotide sequence showing a sequence homology of at least 50% to SEQ ID NO: 3 for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

63. Use of a nucleic acid fragment with the nucleotide sequence shown in SEQ ID NO: 5 or a variant or subsequence of said nucleotide sequence showing a sequence homology of at least 50% to SEQ ID NO: 5 for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

64. Use of a nucleic acid fragment with the nucleotide sequence shown in SEQ ID NO: 7 or a variant or subsequence of

said nucleotide sequence showing a sequence homology of at least 50% to SEQ ID NO: 7 for immunizing a mammal, such as a human, against *Chlamydia pneumoniae*.

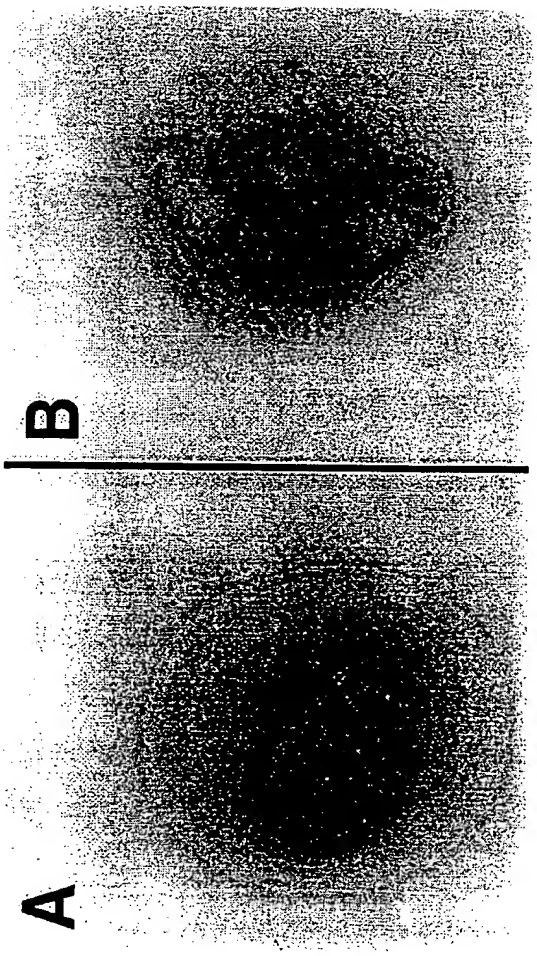


Fig. 1

Fig. 1

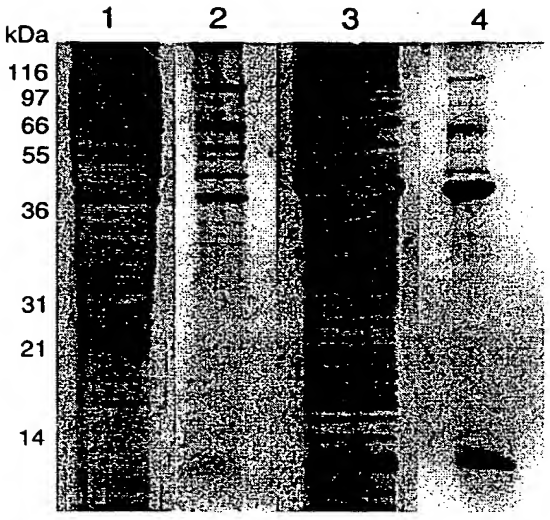


Fig. 2

Fig. 2

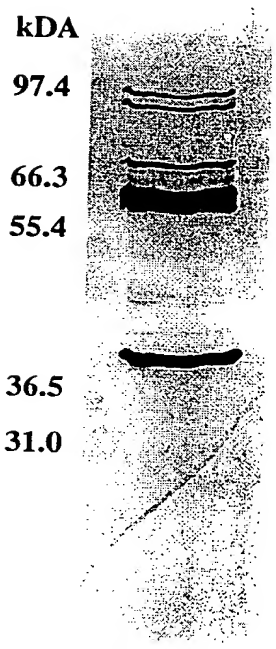


Fig. 3

Fig. 3

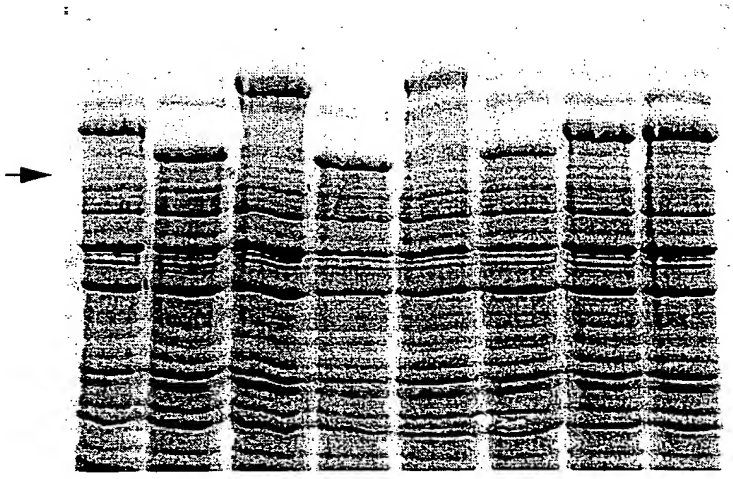


Fig. 4

Fig. 4

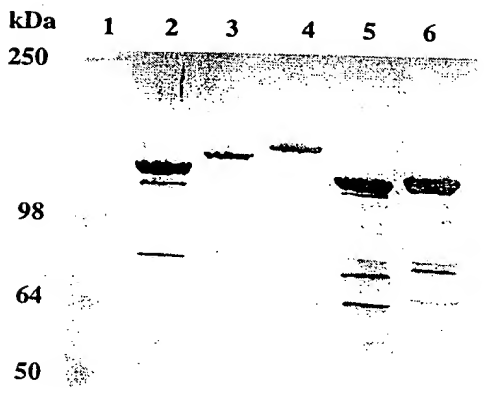
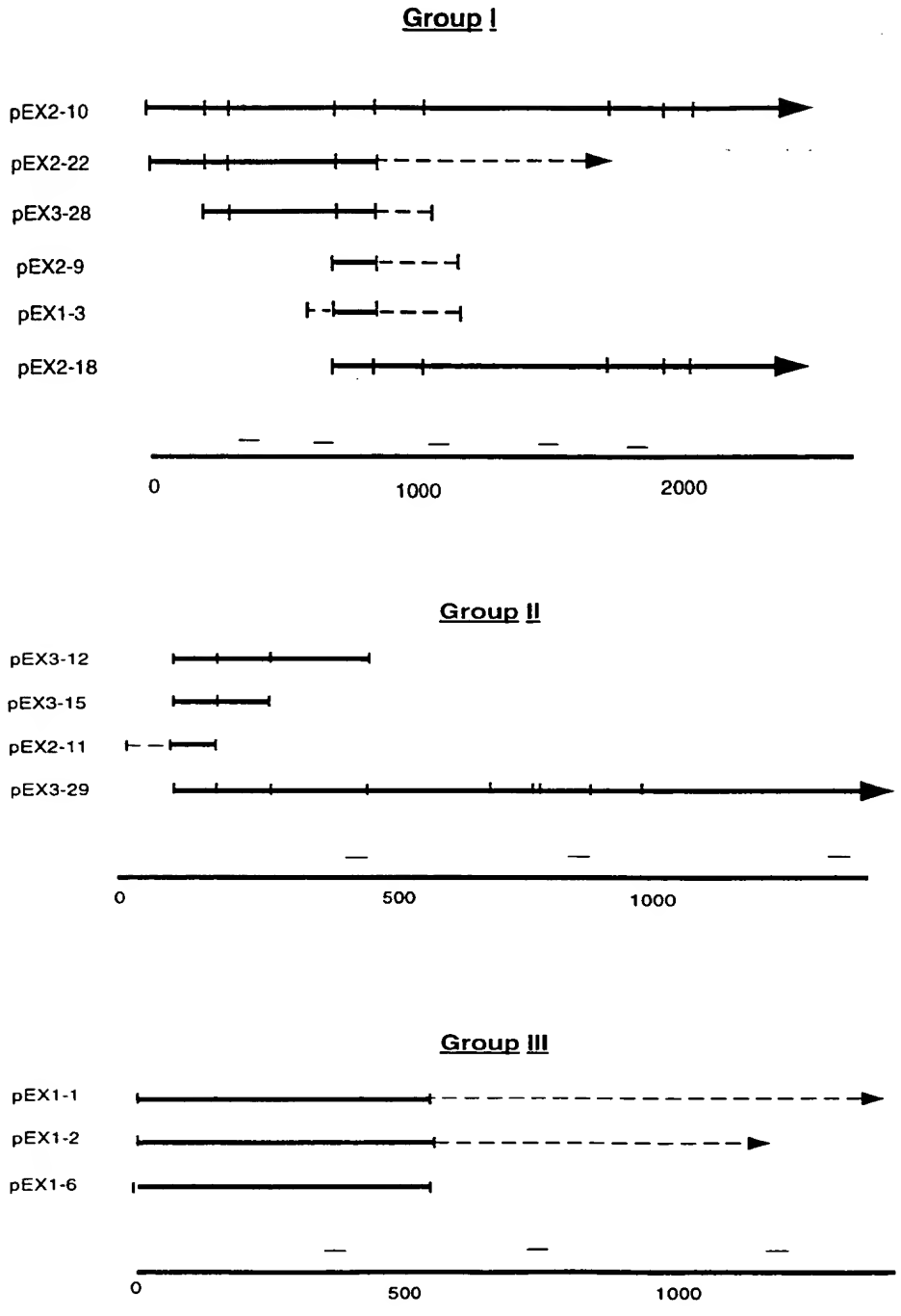


Fig. 5

Fig. 5



Assembled pEX contics.

Fig. 6

Fig. 6

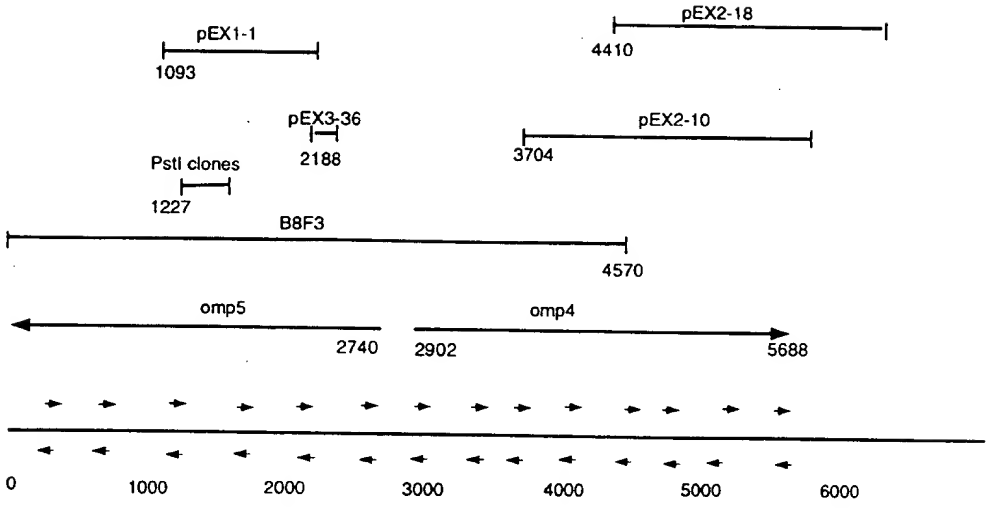


Fig. 7

Fig. 7

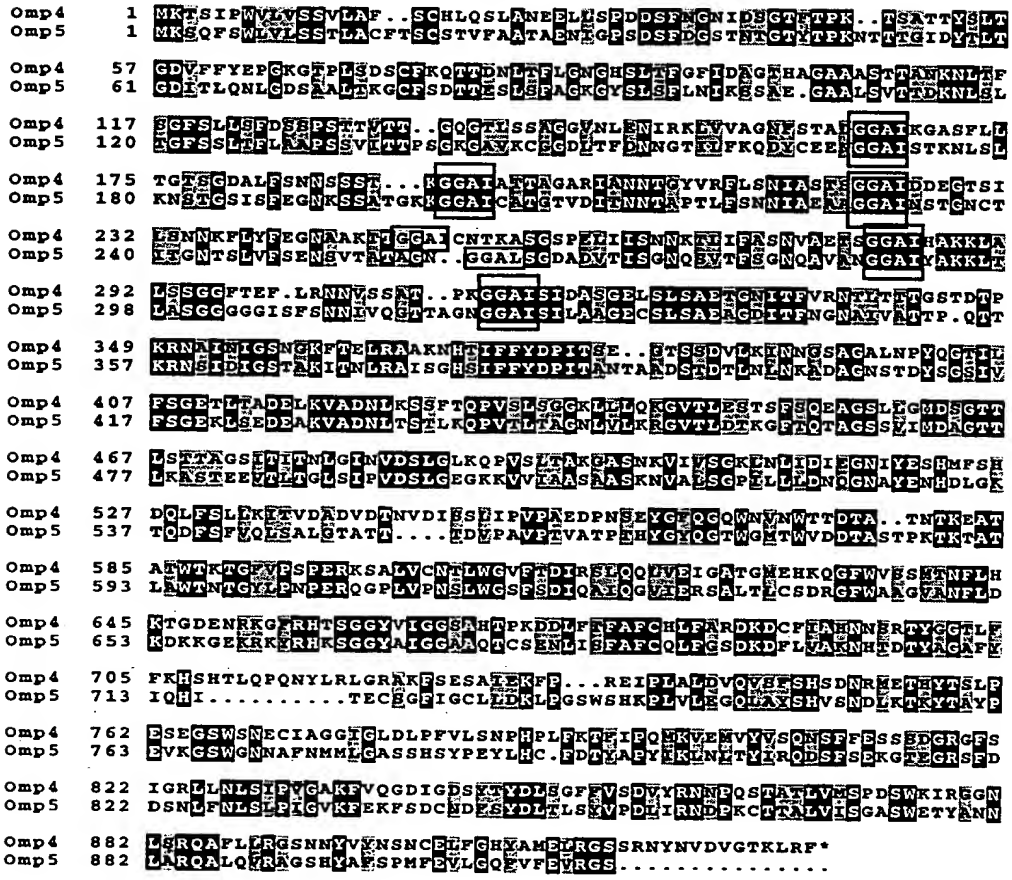


Fig. 8

Fig. 8

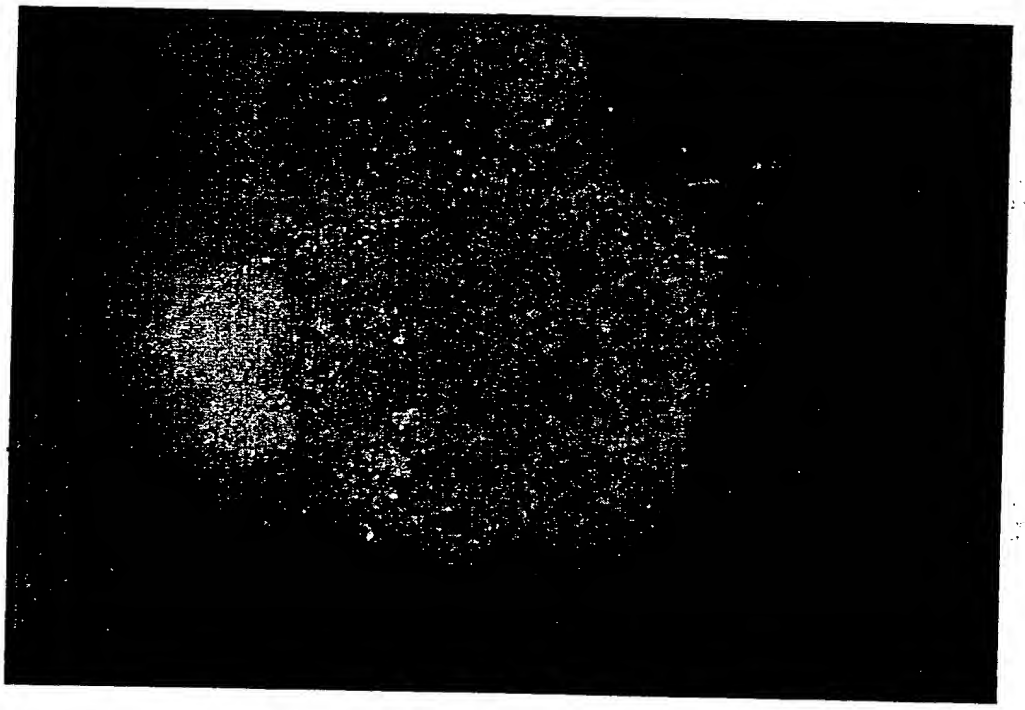


Fig. 9

Fig. 9

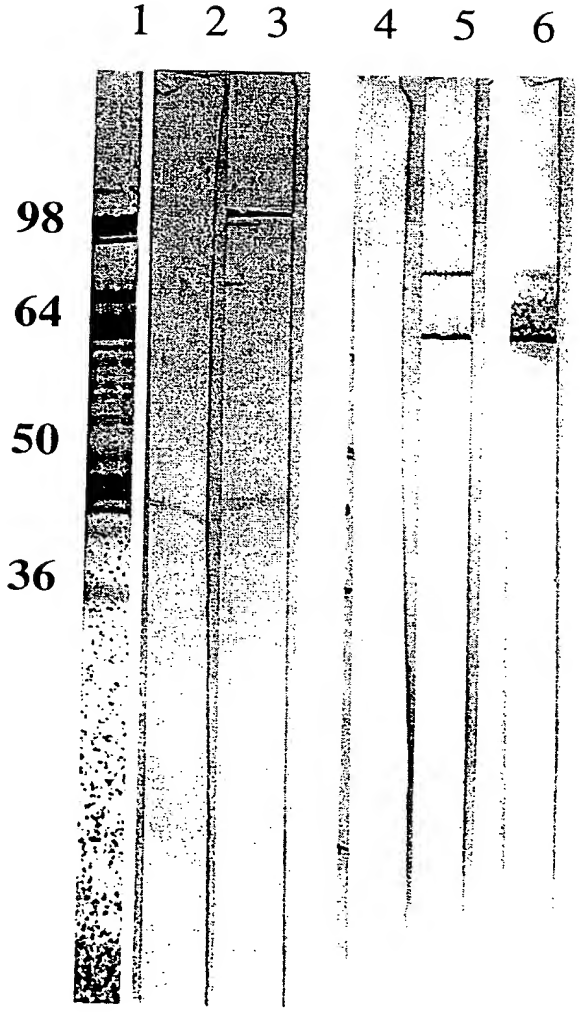


Fig. 10

Fig. 10

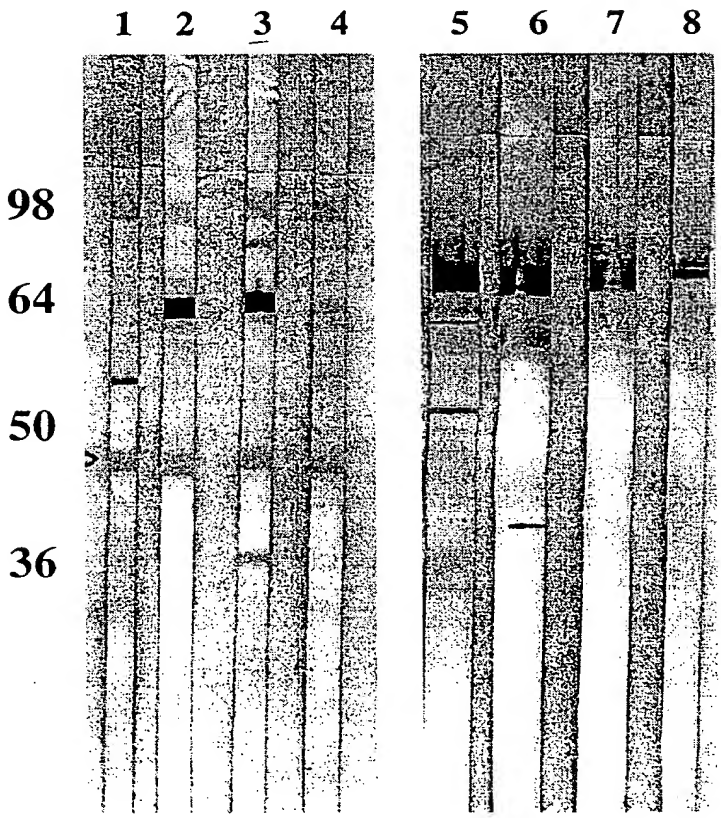
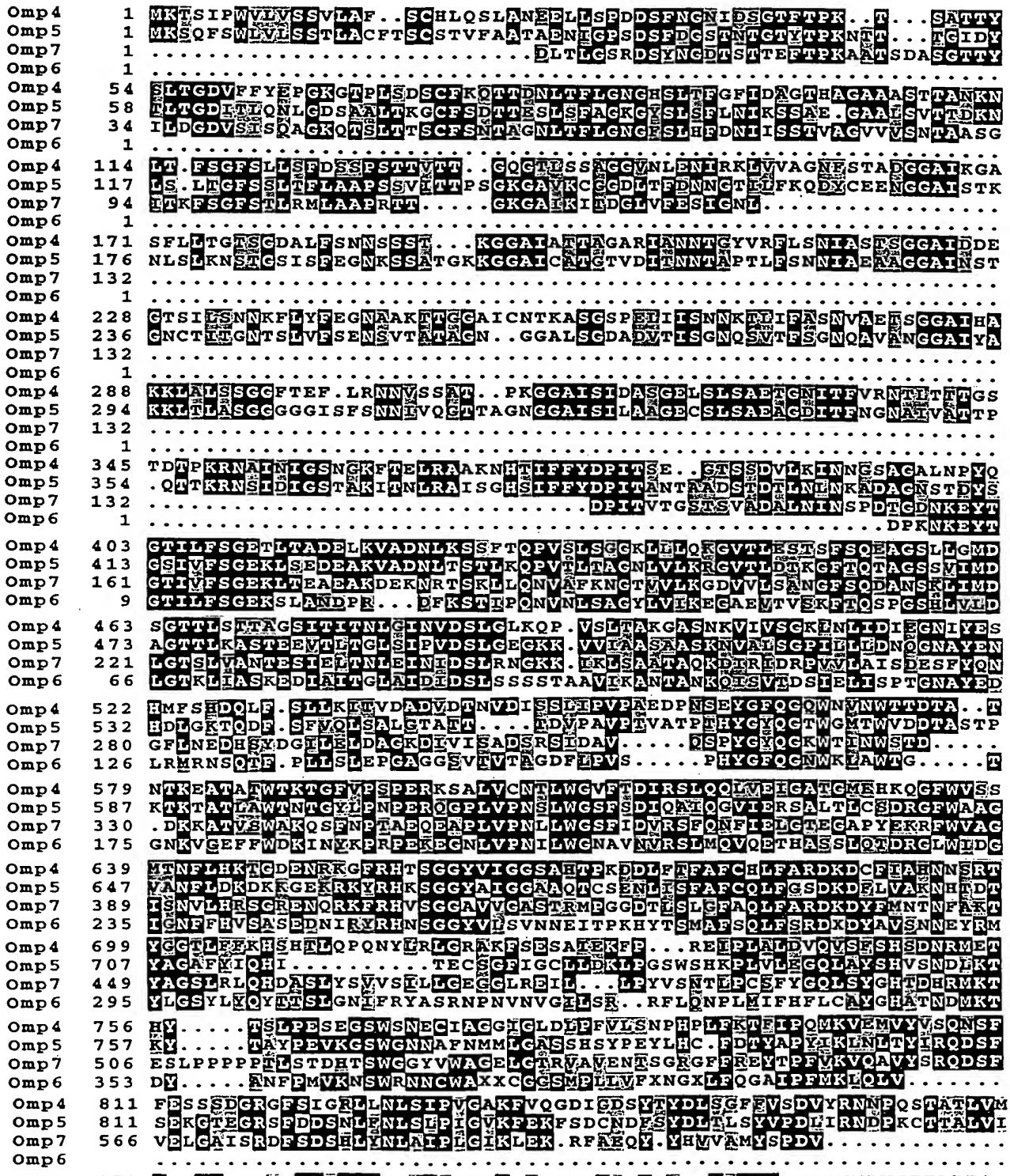


Fig. 11

Fig. 11



BEST AVAILABLE COPY

Fig. 12

Fig. 12